

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

BA

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/12, C07K 14/47, 16/18, C12Q 1/68, G01N 33/577</b>		<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/58061</b> <b>(43) International Publication Date:</b> 23 December 1998 (23.12.98)
<b>(21) International Application Number:</b> PCT/US98/12236 <b>(22) International Filing Date:</b> 18 June 1998 (18.06.98) <b>(30) Priority Data:</b> 08/878,730 19 June 1997 (19.06.97) US 08/878,878 19 June 1997 (19.06.97) US 60/061,641 9 October 1997 (09.10.97) US <b>(71) Applicant:</b> SCHERING CORPORATION [US/US]; 2000 Galloping Hill Road, Kenilworth, NJ 07033-0530 (US). <b>(72) Inventors:</b> FRANZ-BACON, Karin; 5240 Fiore Terrace #J201, San Diego, CA 92122 (US). GORMAN, Daniel, M.; 6371 Central Avenue, Newark, CA 94560 (US). McCLANAHAN, Terrill, K.; 1081 Westchester Drive, Sunnyvale, CA 94087 (US). <b>(74) Agents:</b> THAMPOE, Immac, J. et al.; Schering-Plough Corporation, Patent Dept., K-6-1 1990, 2000 Galloping Hill Road, Kenilworth, NJ 07033-0530 (US).			<b>(81) Designated States:</b> AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CZ, EE, GE, GW, HU, ID, IL, IS, JP, KG, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UZ, VN, YU, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> MAMMALIAN GENES; RELATED REAGENTS			
<b>(57) Abstract</b>  Nucleic acids encoding a new family of small cysteine rich soluble proteins, from a mammal, reagents related thereto, including specific antibodies, and purified proteins are described. Methods of using said reagents and related diagnostic kits are also provided.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakistan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

5

10

## MAMMALIAN GENES; RELATED REAGENTS

## FIELD OF THE INVENTION

The present invention contemplates compositions related to proteins which function in modulating animal physiology, e.g., controlling development, differentiation, trafficking, and physiology of mammalian cells e.g., cells of a mammalian immune system. In particular, it provides purified genes, proteins, antibodies, and related reagents useful, e.g., to regulate activation, development, differentiation, and function of various cell types, particularly primate cells.

25

## BACKGROUND OF THE INVENTION

The circulating component of the mammalian circulatory system comprises various cell types, including red and white blood cells of the erythroid and myeloid cell lineages. See, e.g., Rapaport (1987) Introduction to Hematology (2d ed.) Lippincott, Philadelphia, PA; Jandl (1987) Blood: Textbook of Hematology, Little, Brown and Co., Boston, MA.; and Paul (ed. 1993) Fundamental Immunology (3d ed.) Raven Press, N.Y.

Growth factors, which are typically proteins, influence cellular proliferation and/or differentiation. Usually their effects are mediated through a cell membrane receptor. Some growth factors function via an autocrine signal. The growth factors have varying ranges of specificity of target cells. For example, certain factors play roles in oogenesis, see, e.g., Padgett, et al. (1987) Nature 325:81-84; Ferguson, et al. (1992) Cell

71:451-461; Staehling-Hampton, et al. (1994) Nature 372:783-786; and Panganiban, et al. (1990) Mol. Cell. Biol. 10:2669-2677, in embryogenesis, see, e.g., Xie, et al. (1994) Science 263:1756-1759; Raz, et al. (1993) 5 Genes and Development 7:1937-1948; Brand, et al. (1994) Genes and Development 8:629-639; Goode, et al. (1992) Development 116:177-192; Livneh, et al. (1985) Cell 40:599-607; and Neuman-Silberberg, et al. (1993) Cell 75:164-174; and in morphogenesis, see, e.g., Heberlein, 10 et al. (1993) Cell 75:913-926; Nellen, et al. (1994) Cell 78:225-237; Brummel, et al. (1994) Cell 78:251-261; and Penton, et al. (1994) Cell 78:239-250; of specific cell types or organs.

For some time, it has been known that the mammalian 15 immune response is based on a series of complex cellular interactions, called the "immune network." Recent research has provided new insights into the inner workings of this network. While it remains clear that much of the response does, in fact, revolve around the 20 network-like interactions of lymphocytes, macrophages, granulocytes, and other cells, immunologists now generally hold the opinion that soluble proteins, known as lymphokines, cytokines, or monokines, play a critical role in controlling these cellular interactions. Thus, 25 there is considerable interest in the isolation, characterization, and mechanisms of action of cell modulatory factors, an understanding of which should lead to significant advancements in the diagnosis and therapy of numerous medical abnormalities, e.g., immune system 30 and other disorders.

Lymphokines apparently mediate cellular activities in a variety of ways. They have been shown to support the proliferation, growth, and differentiation of the pluripotential hematopoietic stem cells into vast numbers 35 of progenitors comprising diverse cellular lineages making up a complex immune system. These interactions between the cellular components are necessary for a healthy immune response. These different cellular lineages often respond in a different manner when



lymphokines are administered in conjunction with other agents.

The chemokines are a large and diverse superfamily of proteins. The superfamily is subdivided into three branches, based upon whether the first two cysteines in the classical chemokine motif are adjacent (termed the "C-C" branch) or spaced by an intervening residue ("C-X-C"), or a new branch which lacks two cysteines in the corresponding motif, represented by the chemokines known as lymphotactins. See, e.g., Schall and Bacon (1994) Current Opinion in Immunology 6:865-873; and Bacon and Schall (1996) Int. Arch. Allergy & Immunol. 109:97-109.

There is considerable interest in the isolation, characterization, and mechanisms of action of modulatory factors. Many factors have been identified which influence the differentiation process of precursor cells, or regulate the physiology or migration properties of specific cell types. These observations indicate that other factors exist whose functions in immune function were heretofore unrecognized. These factors provide for biological activities whose spectra of effects may be distinct from known differentiation or activation factors. The absence of knowledge about the structural, biological, and physiological properties of the regulatory factors which regulate cell physiology in vivo prevents the modification of the effects of such factors. Thus, medical conditions where regulation of the development or physiology of relevant cells is required remains unmanageable.

30

#### SUMMARY OF THE INVENTION

The present invention reveals the existence of a new family of Cysteine Rich Soluble Proteins (CRSPs). Their expression suggests a role in immunological function, particularly in inflammatory conditions. It is characterized in various rodent and human embodiments. See also USSN 08/878,878, filed on June 19, 1997, which is incorporated herein by reference. Structural similarity to the defensins, along with expression

levels, suggest a direct antimicrobial function, e.g., microbiostatic or microbiocidal.

Specific embodiments of the present invention provide a composition of matter selected from: a

5 substantially pure or recombinant C2 or C2b polypeptide exhibiting identity over a length of at least 12 amino acids to SEQ ID NO: 2 or 4; a natural sequence C2 of SEQ ID NO: 2 or 4; a fusion protein comprising C2 sequence; a

10 substantially pure or recombinant C18 polypeptide exhibiting identity over a length of at least 12 amino acids to SEQ ID NO: 6; a natural sequence C18 of SEQ ID NO: 6; a fusion protein comprising C18 sequence; a

15 substantially pure or recombinant C19 polypeptide exhibiting sequence identity over a length of at least 12 amino acids to SEQ ID NO: 8 or 10; a natural sequence C19 of SEQ ID NO: 8 or 10; a fusion protein comprising C19

20 sequence; a substantially pure or recombinant C10 polypeptide exhibiting identity over a length of at least 12 amino acids to SEQ ID NO: 12; a natural sequence C10 of SEQ ID NO: 12; or a fusion protein comprising C10 sequence. In various particular forms, it includes a substantially pure or isolated polypeptide comprising a segment exhibiting sequence identity to a corresponding

25 portion of a C2, C2b, C18, C19, or C10 wherein: the identity is over at least 15 amino acids; the identity is over at least 19 amino acids; or the identity is over at least 25 amino acids. Other particular features include such composition, where the C2 or C2b comprises a mature sequence of Table 1; the C18 comprises a mature sequence

30 of Table 2; the C19 comprises a mature sequence of Table 3; the C10 comprises a mature sequence of Table 4; or where the polypeptide: is from a warm blooded animal selected from a mammal, including a rodent or primate; comprises at least one polypeptide segment of SEQ ID NO:

35 2, 4, 6, 8, 10, or 12; exhibits a plurality of portions exhibiting said identity; is a natural allelic variant of C2, C2b, C18, C19, or C10; has a length at least about 30 amino acids; exhibits at least two non-overlapping epitopes which are specific for a mammalian C2, C2b, C18,

C19, or C10; exhibits sequence identity over a length of at least about 20 amino acids to a rodent C2, C2b, C18, or C19 or primate C10; exhibits at least two non-overlapping epitopes which are specific for a rodent C2, C2b, C18, or C19 or primate C10; exhibits sequence identity over a length of at least about 20 amino acids to a primate C2, C2b, C18, C19, or C10; is not glycosylated; has a molecular weight of at least 3 kD; is a synthetic polypeptide; is attached to a solid substrate; is conjugated to another chemical moiety; is a 5-fold or less substitution from natural sequence; or is a deletion or insertion variant from a natural sequence. In certain preferred embodiments, the composition is: a sterile C2 or C2b polypeptide; or the C2 or C2b polypeptide is with a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; a sterile C18 polypeptide; or the C18 polypeptide is with a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; a sterile C19 polypeptide; or the C19 polypeptide is with a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration; a sterile C10 polypeptide; or the C10 polypeptide is with a carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration.

In addition, the present invention comprises a composition of matter selected from: a substantially pure or recombinant C23 protein or peptide exhibiting at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 14; a natural sequence C23 of SEQ ID NO: 14; or a fusion protein comprising C23 sequence. In certain embodiments, the protein comprises a segment exhibiting sequence identity to a corresponding

portion of a C23, wherein: the homology is at least about 90% identity and the portion is at least about 9 amino acids; the homology is at least about 80% identity and the portion is at least about 17 amino acids; or the homology is at least about 70% identity and the portion is at least about 25 amino acids. In other embodiments, the C23 comprises a mature sequence of Table 6; or protein or peptide: is from a warm blooded animal selected from a mammal, including a primate; comprises at least one polypeptide segment of SEQ ID NO: 14; exhibits a plurality of portions exhibiting said identity; is a natural allelic variant of C23; has a length at least about 30 amino acids; exhibits at least two non-overlapping epitopes which are specific for a mammalian C23; exhibits a sequence identity at least about 90% over a length of at least about 20 amino acids to a primate C23; exhibits at least two non-overlapping epitopes which are specific for a primate C23; exhibits a sequence identity at least about 90% over a length of at least about 20 amino acids to a primate C23; is glycosylated; has a molecular weight of at least 7 kD with natural glycosylation; is a synthetic polypeptide; is attached to a solid substrate; is conjugated to another chemical moiety; is a 5-fold or less substitution from natural sequence; or is a deletion or insertion variant from a natural sequence. Preferably, the composition can comprise: a sterile C23 protein or peptide; or the C2 protein or peptide and a carrier, wherein said carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration.

In fusion protein embodiments, the fusion protein comprises: mature protein sequence of Table 1, 2, 3, 4 or 6; a detection or purification tag, including a FLAG, His6, or Ig sequence; or sequence of another cytokine or growth factor protein.

Certain kit embodiments comprise a protein or polypeptide, and: a compartment comprising said protein

or polypeptide; and/or instructions for use or disposal of reagents in said kit.

Various binding compound embodiments are provided, including ones comprising an antigen binding portion from

5 an antibody, which specifically binds to a natural C2, C2b, C18, C19, C10 or C23 protein, wherein: the polypeptide is a rodent or primate protein; the binding compound is an Fv, Fab, or Fab2 fragment; the binding compound is conjugated to another chemical moiety; or the

10 antibody: is raised against a peptide sequence of a mature polypeptide of Table 1, 2, 3, 4 or 6; is raised against a mature C2, C2b, C18, C19, C10 or C23; is raised to a purified C2, C2b, C18, C19, C10 or C23; is immunoselected; is a polyclonal antibody; binds to a

15 denatured C2, C2b, C18, C19, C10 or C23; exhibits a Kd to antigen of at least 30  $\mu$ M; is attached to a solid substrate; including a bead or plastic membrane; is in a sterile composition; or is detectably labeled, including a radioactive or fluorescent label. Other kit

20 embodiments include a kit comprising such binding compound, and: a compartment comprising said binding compound; and/or instructions for use or disposal of reagents in said kit. Preferably, the kit is capable of making a qualitative or quantitative analysis.

25 Methods are provided, e.g., of: A) making such an antibody, comprising immunizing an immune system with an immunogenic amount of: a rodent C2 polypeptide; a rodent C2b polypeptide; a rodent C18 polypeptide; a rodent C19 polypeptide; a primate C10 polypeptide; or a primate C23

30 polypeptide: thereby resulting in production of such antibody; or B) producing an antigen:antibody complex, comprising contacting: a rodent C2 polypeptide with such an antibody; a rodent C2b polypeptide with such an antibody; a rodent C18 polypeptide with such an antibody;

35 a rodent C19 polypeptide with such an antibody; a primate C10; or a primate C23 polypeptide with such an antibody; thereby allowing such complex to form.

Other embodiments include compositions comprising: a sterile binding compound, or the binding compound and a

carrier, wherein the carrier is: an aqueous compound, including water, saline, and/or buffer; and/or formulated for oral, rectal, nasal, topical, or parenteral administration.

5 Nucleic acid embodiments include an isolated or recombinant nucleic acid encoding a CRSP polypeptide or fusion protein, wherein: the C family protein is from a mammal, including a rodent or primate; or the nucleic acid: encodes an antigenic peptide sequence of Table 1,  
10 2, 3, 4 or 6; encodes a plurality of antigenic peptide sequences of Table 1, 2, 3, 4 or 6; exhibits identity to a natural cDNA encoding said segment; is an expression vector; further comprises an origin of replication; is from a natural source; comprises a detectable label;  
15 comprises synthetic nucleotide sequence; is less than 6 kb, preferably less than 3 kb; is from a mammal, including a rodent; comprises a natural full length coding sequence; is a hybridization probe for a gene encoding said C family protein; or is a PCR primer, PCR  
20 product, or mutagenesis primer. The invention also provides a cell or tissue comprising such a recombinant nucleic acid, particularly where the cell is: a prokaryotic cell; a eukaryotic cell; a bacterial cell; a yeast cell; an insect cell; a mammalian cell; a mouse  
25 cell; a primate cell; or a human cell. Kits containing such nucleic acids may also include a compartment comprising the nucleic acid; a compartment further comprising a C2, C2b, C18, C19, C10 or C23 protein or polypeptide; and/or instructions for use or disposal of  
30 reagents in the kit. Preferably, the kit is capable of making a qualitative or quantitative analysis.

Other methods are provided, e.g., of: A) making a polypeptide, comprising expressing such a nucleic acid, thereby producing such polypeptide; or B) making a duplex  
35 nucleic acid, comprising contacting such nucleic acid with a hybridizing nucleic acid, thereby allowing such duplex to form.

Other nucleic acid embodiments include those which: hybridize under wash conditions of 30° C and less than 2M

salt to SEQ ID NO: 1; hybridize under wash conditions of 30° C and less than 2 M salt to SEQ ID NO: 3; hybridize under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 5; hybridize under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 7; hybridize under wash conditions of 30° C and less than 2 M salt to SEQ ID NO: 9; hybridize under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 11; exhibit at least about 85% identity over a stretch of at least about 30 nucleotides to a rodent C2; exhibit at least about 85% identity over a stretch of at least about 30 nucleotides to a rodent C2b; exhibit at least about 85% identity over a stretch of at least about 30 nucleotides to a rodent C18; exhibit at least about 85% identity over a stretch of at least about 30 nucleotides to a rodent C19; or exhibit at least about 85% identity over a stretch of at least about 30 nucleotides to a primate C10. Other embodiments include those wherein: the wash conditions are at 45° C and/or 500 mM salt; the wash conditions are at 55° C and/or 150 mM salt; the identity is at least 90% and/or the stretch is at least 55 nucleotides; or the identity is at least 95% and/or the stretch is at least 75 nucleotides.

Additional nucleic acid embodiments include ones which: hybridize under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 13; where the wash conditions are at 45° C and/or 500 mM salt; where the wash conditions are at 55° C and/or 150 mM salt; or exhibit at least about 85% identity over a stretch of at least about 30 nucleotides to a primate C23; or the identity is at least 90% and/or the stretch is at least 55 nucleotides; or where the identity is at least 95% and/or the stretch is at least 75 nucleotides.

The invention also provides a method of modulating physiology or development of a cell or tissue culture cells comprising contacting said cell with an agonist or antagonist of a C2, C2b, C18, C19, C10 or C23. This may have a direct or indirect effect on other cells.

## DETAILED DESCRIPTION

All references cited herein are incorporated herein  
5 by reference to the same extent as if each individual  
publication or patent application was specifically and  
individually indicated to be incorporated by reference.

## 10 I. General

The present invention provides DNA sequences  
encoding soluble mammalian proteins which exhibit  
structural properties or motifs characteristic of a small  
soluble protein, e.g., a defensin, growth factor,  
15 cytokine, or chemokine. Because the proteins are  
cysteine rich, and the cysteine motifs appear conserved,  
these proteins are referred to as Cysteine Rich Soluble  
Proteins (CRSPs). The tissue expression distribution of  
these proteins correlate with significant medical  
20 conditions.

For reviews on the defensins, see, e.g., Ganz, et  
al. (1998) Current Op. in Immunol. 10:41-44; Hancock, et  
al. (1998) Trends Biotech. 16:82-88; Lehrer, et al.  
(1996) Ann. NY Acad. Sci. 797:228-239; White, et al.  
25 (1995) Current Op. Struct. Biol. 5:521-527; Ganz, et al.  
(1995) Pharmacol. Ther. 66:191-205; Harwig, et al. (1994)  
Methods in Enzymology 236:160-172; and Lehrer, et al.  
(1993) Ann. Rev. Immunol. 11:105-128. The defensins  
exhibit significant sequence similarity to the carboxy  
30 terminus of the CRSPs. Processing of the defensins from  
inactive precursors follows a known pathway. The amounts  
of protein expressed and the exon structure of the  
defensins seems to match with that of the CRSPs.

For a review of the cytokines, see, e.g., Thompson  
35 (1994) The Cytokine Handbook 2d ed., Academic Press, San  
Diego; and Aggarwal and Gutterman (1992) Human Cytokines:  
Handbook for Basic and Clinical Research, Blackwell Pub.,  
Oxford. Many specific sequences and references are  
available, e.g., from the GenBank, and references



providing gene and/or cytokine amino acid sequence. Many receptor sequences are also available from GenBank. See also Howard, et al. (1993) in Paul (ed. 1993) Fundamental Immunology (3d ed.) Raven Press, NY.

5       For reviews of the chemokine family, see, e.g., Lodi, et al. (1994) Science 263:1762-1767; Gronenborn and Clore (1991) Protein Engineering 4:263-269; Miller and Kranger (1992) Proc. Nat'l Acad. Sci. USA 89:2950-2954; Matsushima and Oppenheim (1989) Cytokine 1:2-13; Stoeckle  
10   and Baker (1990) New Biol. 2:313-323; Oppenheim, et al. (1991) Ann. Rev. Immunol. 9:617-648; Schall (1991) Cytokine 3:165-183; and The Cytokine Handbook Academic Press, NY. The proteins described herein are designated  
15   Cysteine Rich Soluble Proteins because they were initially recognized as a class of soluble proteins. These proteins share a highly conserved pattern of cysteine motifs, e.g., structural motifs, distinct from the other known groups of soluble protein molecules.

      The best characterized embodiments of this family of  
20   proteins were initially discovered from mouse sequence sources (see Tables 1, 2, 3, and 4) as well as from a human sequence source (Table 6). The descriptions below are directed, for exemplary purposes, to the expressly provided embodiments, but are likewise applicable to  
25   related embodiments from other, e.g., natural, sources. These sources should include various vertebrates, typically warm blooded animals, e.g., birds and mammals, particularly domestic animals, and primates.

Table 1: Mouse nucleic acid sequence (SEQ ID NO: 1) and corresponding amino acid sequence (SEQ ID NO: 2) of cysteine rich soluble protein 2 (C2). Coding sequence begins at about nucleotide 32 and ends at about nucleotide 364 (end of last coding codon before the termination). Experimental evidence suggests mature protein begins as indicated with DET..., so the signal peptide runs from amino acid -23 (Met) to about predicted Thr (-1). Helical structures run from mature protein residues 2 (Thr) to 16 (Leu).  $\beta$  sheet structures run from about 30 (Leu) to 36 (Lys); 41 (Trp) to 43 (Ser); 49 (Thr) to 52 (Gly); 62 (Trp) to 65 (Gln); 70 (Cys) to 74 (Cys); and 85 (Cys) to 88 (Ser). Structure is based upon use of PHD program: accessed by <http://www.embl-heidelberg.de/predictprotein/>; or by the DSC program: accessed by <http://bonsai.lif.icnet.uk/bmm/dsc/>. Intron appears to be between about nucleotides 158-159 and 239-240.

```

5  ATTCTGCCCC AGGATGCCAA CTTTGAATAG G ATG AAG ACT ACA ACT TGT TCC
    52
    Met Lys Thr Thr Thr Cys Ser
    -23          -20

20  CTT CTC ATC TGC ATC TCC CTG CTC CAG CTG ATG GTC CCA GTG AAT ACT
    100
    Leu Leu Ile Cys Ile Ser Leu Leu Gln Leu Met Val Pro Val Asn Thr
    25  -15          -10          -5

    GAT GAG ACC ATA GAG ATT ATC GTG GAG AAT AAG GTC AAG GAA CTT CTT
    148
    Asp Glu Thr Ile Glu Ile Ile Val Glu Asn Lys Val Lys Glu Leu Leu
    30  1          5          10          15

    GCC AAT CCA GCT AAC TAT CCC TCC ACT GTA ACG AAG ACT CTC TCT TGC
    196
    Ala Asn Pro Ala Asn Tyr Pro Ser Thr Val Thr Lys Thr Leu Ser Cys
    35  20          25          30

    ACT AGT GTC AAG ACT ATG AAC AGA TGG GCC TCC TGC CCT GCT GGG ATG
    244
    Thr Ser Val Lys Thr Met Asn Arg Trp Ala Ser Cys Pro Ala Gly Met
    40  35          40          45

    ACT GCT ACT GGG TGT GCT TGT GGC TTT GCC TGT GGA TCT TGG GAG ATC
    292
    Thr Ala Thr Gly Cys Ala Cys Gly Phe Ala Cys Gly Ser Trp Glu Ile
    45  50          55          60

    CAG AGT GGA GAT ACT TGC AAC TGC CTG TGC TTA CTC GTT GAC TGG ACC
    340
    Gln Ser Gly Asp Thr Cys Asn Cys Leu Cys Leu Leu Val Asp Trp Thr
    50  65          70          75          80

    ACT GCC CGC TGC TGC CAA CTG TCC TAAGAATGAA GAGGTGGAGA ACCCAGCTTT
    394
    Thr Ala Arg Cys Cys Gln Leu Ser
    55  85

```

GATATGATGA ATCTAACAAA AACTGCAGTC TCAATTGGA AATCTGACTC ATGTGCCTTT  
454

AAATGTGTTC ATATTGCCCA TTTACCCTGC TTCTTGAAAT GCTTCTTGAA AAATAAGAC  
5 514

AAATTTGCAT GTG  
527

10 Table 1 (continued) A closely related gene, mouse C2b, has also  
been identified (SEQ ID NO: 3 and 4). A predicted signal sequence  
is indicated; experimental determination suggests a blocked N-  
terminus; genomic analysis indicates introns between about  
nucleotides 196-197 and 277-278:

15 AGCATCTCAT CTGGCCAGGT CCTGGAACCT TTCCTGAGAT TCTGCCCTAG GATGCTGACT  
60

TTCAACAAG ATG AAG ACT ACA ACT TGT TCC CTT CTC ATC TGC ATC TCC  
108

20 Met Lys Thr Thr Thr Cys Ser Leu Leu Ile Cys Ile Ser  
-23 -20 -15  
CTT CTC CAG CTG ATG GTC CCA GTG AAT ACT GAG GGG ACC TTA GAA TCT  
156

25 Leu Leu Gln Leu Met Val Pro Val Asn Thr Glu Gly Thr Leu Glu Ser  
-10 -5 1 5

ATT GTG GAG AAA AAG GTC AAG GAA CTT CTT GCC AAT CGA GAT GAC TGT  
204

30 Ile Val Glu Lys Lys Val Lys Glu Leu Leu Ala Asn Arg Asp Asp Cys  
10 15 20

CCC TCC ACT GTA ACA AAG ACT TTC TCC TGT ACT AGT ATC ACG GCT TCA  
252

35 Pro Ser Thr Val Thr Lys Thr Phe Ser Cys Thr Ser Ile Thr Ala Ser  
25 30 35

GGC AGA CTG GCC TCC TGT CCT TCT GGA ATG ACT GTC ACT GGT TGT GCT  
300

40 Gly Arg Leu Ala Ser Cys Pro Ser Gly Met Thr Val Thr Gly Cys Ala  
40 45 50

TGT GGC TAT GGC TGT GGA TCT TGG GAT ATC CGG GAT GGA AAT ACT TGC  
348

45 Cys Gly Tyr Gly Cys Gly Ser Trp Asp Ile Arg Asp Gly Asn Thr Cys  
55 60 65 70

CAC TGT CAG TGC TCA ACA ATG GAC TGG GCC ACC GCC CGT TGC TGC CAA  
396

50 His Cys Gln Cys Ser Thr Met Asp Trp Ala Thr Ala Arg Cys Cys Gln  
75 80 85

CTG GCC TAAGAATGAG GAAGCTGAGA ACCTAGCTTT GAAATGAAGA CTATAACAAA  
452

55 Leu Ala

AGCACAATCC CAACTTGGA ACCTGGCTCA TATCCCATTTG ATGAATTCAT ATTGTCCATT  
512

60 AGCCCTGCTT CTTGAAAAA ATAAAGACAA ATTTGCACGT GTCTGTAAAA AAAAAAAAAA  
572

AA  
574

Table 2: Mouse nucleic acid sequence (SEQ ID NO: 5) and corresponding amino acid sequence (SEQ ID NO: 6) of cysteine rich soluble protein embodiment designated C18. Coding sequence begins at about nucleotide 103 and ends at about nucleotide 417 (end of last coding codon before the termination). Putative PSORT signal peptide runs from amino acid -19 (Met) to predicted -1 (Val), so mature protein may start at Prol. SignalP predicts that mature sequence begins at Gln5. Signal may be slightly longer or shorter, depending upon cell type, etc. Introns between about nucleotides 230-231 and 292-293. Mature protein sequence should lack signal sequence.

```

5  CCTGAGCTTT CTGGAGAGTG AATCTGCTCT TAGGGGAAAA GCTCTTCCTT TTCCTTCTCC
10 60
15  AAAAAGCTAG AACTGAGCTC CAGGAGGCTG ACTTTCTACA GC ATG AAG CCT ACA
114
Met Lys Pro Thr
-19
20  CTG TGT TTC CTT TTC ATC CTC GTC TCC CTT TTC CCA CTG ATA GTC CCA
162
Leu Cys Phe Leu Phe Ile Leu Val Ser Leu Phe Pro Leu Ile Val Pro
-15 -10 -5 1
25  GGG AAC GCG CAA TGC TCC TTT GAG TCT TTG GTG GAT CAA AGG ATC AAG
210
Gly Asn Ala Gln Cys Ser Phe Glu Ser Leu Val Asp Gln Arg Ile Lys
5 10 15
30  GAA GCT CTC AGT CGT CAA GAG CCT AAG ACG ATC TCC TGC ACT AGT GTC
258
Glu Ala Leu Ser Arg Gln Glu Pro Lys Thr Ile Ser Cys Thr Ser Val
20 25 30
35  ACG TCT TCT GGC AGA CTG GCC TCC TGT CCT GCT GGG ATG GTT GTC ACT
306
Thr Ser Ser Gly Arg Leu Ala Ser Cys Pro Ala Gly Met Val Val Thr
35 40 45
40  GGA TGT GCT TGT GGC TAT GGC TGT GGA TCG TGG GAT ATC CGG AAT GGA
354
Gly Cys Ala Cys Gly Tyr Gly Cys Gly Ser Trp Asp Ile Arg Asn Gly
50 55 60 65
45  AAT ACT TGC CAC TGC CAG TGC TCA GTC ATG GAC TGG GCC TCT GCC CGC
402
Asn Thr Cys His Cys Gln Cys Ser Val Met Asp Trp Ala Ser Ala Arg
70 75 80
50  TGC TGC CGA ATG GCT TAAGAATGAG GAGGTTGAGA AACCAATTTT AAAATGATGA
457
Cys Cys Arg Met Ala
85
55  CCATAATGAA ACCACGGTCT CGACCAGGAA ACCTGACTCA TTGTCTTCAT ATTACTAAAT
517
60  AATTCTTCTT GAATAATAAA GGCAGACCTG TACCTTT
554

```

Table 3: A mouse nucleic acid sequence (SEQ ID NO: 7) and corresponding amino acid sequence (SEQ ID NO: 8) of cysteine rich soluble protein embodiment designated C19. Coding sequence begins at about nucleotide 64 and ends at about nucleotide 405 (end of last coding codon before the termination). Sequencing of mature protein indicates signal peptide runs from amino acid -20 (Met) to predicted -1 (Gly), as indicated. Introns between about nucleotides 193-194, and 271-272.

10	GACAGGAGCT AATACCCAGA ACTGAGTTGT GTCCTGCTAA GTCCTCTGCC ACGTACCCAC 60
15	GGG ATG AAG AAC CTT TCA TTT CCC CTC CTT TTC CTT TTC TTC CTT GTC 108 Met Lys Asn Leu Ser Phe Pro Leu Leu Phe Leu Phe Phe Leu Val -20 -15 -10
20	CCT GAA CTG CTG GGC TCC AGC ATG CCA CTG TGT CCC ATC GAT GAA GCC 156 Pro Glu Leu Leu Gly Ser Ser Met Pro Leu Cys Pro Ile Asp Glu Ala -5 1 5 10
25	ATC GAC AAG AAG ATC AAA CAA GAC TTC AAC TCC CTG TTT CCA AAT GCA 204 Ile Asp Lys Lys Ile Lys Gln Asp Phe Asn Ser Leu Phe Pro Asn Ala 15 20 25
30	ATA AAG AAC ATT GGC TTA AAT TGC TGG ACA GTC TCC TCC AGA GGG AAG 252 Ile Lys Asn Ile Gly Leu Asn Cys Trp Thr Val Ser Ser Arg Gly Lys 30 35 40
35	TTG GCC TCC TGC CCA GAA GGC ACA GCA GTC TTG AGC TGC TCC TGT GGC 300 Leu Ala Ser Cys Pro Glu Gly Thr Ala Val Leu Ser Cys Ser Cys Gly 45 50 55
40	TCT GCC TGT GGC TCG TGG GAC ATT CGT GAA GAA AAA GTG TGT CAC TGC 348 Ser Ala Cys Gly Ser Trp Asp Ile Arg Glu Glu Lys Val Cys His Cys 60 65 70 75
45	CAG TGT GCA AGG ATA GAC TGG ACA GCA GCC CGC TGC TGT AAG CTG CAG 396 Gln Cys Ala Arg Ile Asp Trp Thr Ala Ala Arg Cys Cys Lys Leu Gln 80 85 90
50	GTC GCT TCC TGATGTCGGG GAAGTGAGCG TGGTTTCCAG CACAGCCACC 445 Val Ala Ser
55	CGTTCCTGTA GCTCCAGAGA TGTCTGATGT CCTCCGGTCT CTACAGGCAC CTGCACTCAC 505 GTGCGCGAAT CCACACACAA GCACACATAC TTAAAAATAA AACAAAACAG GCTGG 560

Table 3 (continued) A rat gene which appears to be a C19 counterpart has been identified (SEQ ID NO: 9 and 10). Introns between about nucleotides 159-160, and 236-237. A predicted signal sequence is indicated:

5 CTGAGCTCTC TGCCACGTAC TTAACAGG ATG AAG AAC CTT TCA TTT CTC CTC  
52  
Met Lys Asn Leu Ser Phe Leu Leu  
10 -17 -15 -10

CTT TTC CTT TTC TTC CTT GTC CTG GGG CTG CTG GGC CCC AGC ATG TCA  
100  
Leu Phe Leu Phe Phe Leu Val Leu Gly Leu Leu Gly Pro Ser Met Ser  
15 -5 1 5

CTG TGT CCC ATG GAT GAA GCC ATC AGC AAG AAG ATC AAT CAA GAC TTC  
148  
Leu Cys Pro Met Asp Glu Ala Ile Ser Lys Lys Ile Asn Gln Asp Phe  
20 10 15 20

AGC TCC CTA CTG CCA GCT GCA ATG AAG AAC ACT GTC CTA CAT TGC TGG  
196  
Ser Ser Leu Leu Pro Ala Ala Met Lys Asn Thr Val Leu His Cys Trp  
25 25 30 35

TCA GTC TCC TCC AGA GGG AGG CTG GCC TCC TGC CCA GAA GGC ACA ACC  
244  
Ser Val Ser Ser Arg Gly Arg Leu Ala Ser Cys Pro Glu Gly Thr Thr  
30 40 45 50 55

GTC ACT AGC TGC TCC TGT GGC TCT GGC TGT GGC TCA TGG GAC GTC CGT  
292  
Val Thr Ser Cys Ser Cys Gly Ser Gly Cys Gly Ser Trp Asp Val Arg  
35 60 65 70

GAG GAT ACA ATG TGT CAC TGC CAG TGC GGA AGC ATA GAC TGG ACA GCG  
340  
Glu Asp Thr Met Cys His Cys Gln Cys Gly Ser Ile Asp Trp Thr Ala  
40 75 80 85

GCC CGC TGC TGT ACC CTG CGG GTT GGT TCC TGAGGACGGT TGATTGAGAA  
390  
Ala Arg Cys Cys Thr Leu Arg Val Gly Ser  
45 90 95

CTGAGCTTGC CCTCCGAGTG CTGCCGAGGG ATGAGCTTGC CCACCATGCC CTGCAGAGGA  
450

GGGATGGGGA TGGGGAGAGC GCAGGGGGCA GGAAACGAGA TGAGGGTTTG GAAATACACA  
50 510

ATGGGATGAT GGTGGTGATA AAGATGCACG GTAAAGTGA AAAAAAAAAA AAAAAAAAAA  
55 570

AA  
572

Table 4: A human cysteine rich soluble protein embodiment designated C10 (SEQ ID NO: 11 and 12). Introns likely between about nucleotides 234-235, and 315-316. Experimentally, N-terminus appears to be blocked, a predicted signal sequence is indicated:

```

5  GGCACGAGGC CACGTTGTCT TCTTTCCTTC ACCACCACCC AGGAGCTCAG AGATCTAAGC
   60

10 TGCTTTCCAT CTTTCTCCC AGCCCCAGGA CACTGACTCT GTACAGG ATG GGG CCG
   116
                                     Met Gly Pro
                                     -20

15 TCC TCT TGC CTC CTT CTC ATC CTA ATC CCC CTT CTC CAG CTG ATC AAC
   164
   Ser Ser Cys Leu Leu Leu Ile Leu Ile Pro Leu Leu Gln Leu Ile Asn
       -15                      -10                      -5

20 CCG GGG AGT ACT CAG TGT TCC TTA GAC TCC GTT ATG GAT AAG AAG ATC
   212
   Pro Gly Ser Thr Gln Cys Ser Leu Asp Ser Val Met Asp Lys Lys Ile
       1                      5                      10                      15

25 AAG GAT GTT CTC AAC AGT CTA GAG TAC AGT CCC TCT CCT ATA AGC AAG
   260
   Lys Asp Val Leu Asn Ser Leu Glu Tyr Ser Pro Ser Pro Ile Ser Lys
           20                      25                      30

30 AAG CTC TCG TGT GCT AGT GTC AAA AGC CAA GGC AGA CCG TCC TCC TGC
   308
   Lys Leu Ser Cys Ala Ser Val Lys Ser Gln Gly Arg Pro Ser Ser Cys
           35                      40                      45

35 CCT GCT GGG ATG GCT GTC ACT GGC TGT GCT TGT GGC TAT GGC TGT GGT
   356
   Pro Ala Gly Met Ala Val Thr Gly Cys Ala Cys Gly Tyr Gly Cys Gly
           50                      55                      60

40 TCG TGG GAT GTT CAG CTG GAA ACC ACC TGC CAC TGC CAG TGC AGT GTG
   404
   Ser Trp Asp Val Gln Leu Glu Thr Thr Cys His Cys Gln Cys Ser Val
       65                      70                      75

45 GTG GAC TGG ACC ACT GCC CGC TGC TGC CAC CTG ACC TGACAGGGAG
   450
   Val Asp Trp Thr Thr Ala Arg Cys Cys His Leu Thr
       80                      85                      90

50 GAGGCTGAGA ACTCAGTTTT GTGACCATGA CAGTAATGAA ACCAGGGTCC CAACCAAGAA
   510

   ATCTAACTCA AACGTCCCAC TTCATTTGTT CCATTCCTGA TTCTTGGGTA ATAAAGACAA
   570

55 ACTTTGTACC TCAAAAAAAAA AAAAAAAAAA AAA
   603

```

Table 5: Comparison of members of CRSP family of proteins. Experimentally determined (mC2, hC23, and mC19) or predicted signal sequences are underlined, though the processing may depend upon the cell type, and may vary by a few amino acids in either direction. Note the conserved cysteines which correspond to the mouse C2 mature residues (2), 32, 44, 53, 55, 59, 70, 72, 74, 84, and 85. An alpha-helical stretch corresponds roughly to C2 residues 2-16, and corresponding structures provided in Table 1.

10

hc10 MGPSSCLLLILIP-LLOLINPGSTQCSLDSVMDKKIKDVLNSLEYSPSPI  
 mC18 MKP-TLCFLFILVSLFPLITVPGNAQCSFESLVDQRIKEALS-----E  
 15 mC2 MKTTTCSLLICIS-LLOLMVPVNTDETIEIIVENKVKELLANPANYPSTV  
 mC2b MKTTTCSLLICIS-LLOLMVPVNTDEGTLESIVEKKVKELLANRDCPSTV  
 hC23 MKALCLLLLP---VLGLLVSSKTLCSMEEAINERIQEVAGSLIFR-AIS  
 mC19 MKNLSFPLLEFLFVLPELLGSSMPLCPIDEAIDKKIKQDFNSLFPN-AIK  
 rC19 MKNLSFLLLEFLFVLGLLGPMSLCPMDEAISKKINQDFSSLLPA-AMK

20

hc10 SKKLSCASVKSQGRPSSCPAGMAVTGCACGYGCGSWDVQLETTCHCQCSV  
 mC18 PKTISCTSVTSSGRLASCAGMVVTGCACGYGCGSWDIRNGNTCHCQCSV  
 mC2 TKTLSCTSVKTMNRWASCPAGMTATGCACGFACGSWEIQSGDTCNCLCLL  
 25 mC2b TKTFSCTSITASGRLASCPSGMTVTGACGYGCGSWDIRDGNTHCQCSV  
 hC23 SIGLEQSVTSRGLATCPRGFAVTGCTCGSACGSWDVRAETTHCQCSV  
 mC19 NIGLNCWTVSSRGLASCPEGTAVLSCSCGSACGSWDIREKVVCHCQCAR  
 rC19 NTVLHCWSVSSRGLASCPEGTTVTSCSCGSACGSWDVREDTMCHCQCSV

30

hc10 VDWTARCCHLT  
 mC18 MDWASARCCMA  
 mC2 VDWTARCCQLS  
 mC2b MDWATARCCQLA  
 35 hC23 MDWTGARCCRVQP  
 mC19 IDWTAARCCKLQVAS  
 rC19 IDWTAARCCCLRVS

40 Table 6: Human nucleic acid sequence (SEQ ID NO: 13) and  
 corresponding amino acid sequence (SEQ ID NO: 14) of cysteine rich  
 soluble protein 23 (C23). Coding sequence begins at about  
 nucleotide 47 and ends at about nucleotide 370 (end of last coding  
 codon before the termination). Putative signal peptide runs from  
 45 amino acid -18 (Met) to about predicted -1 (Ser). Mature sequence  
 has signal sequence removed. Helical structures run from mature  
 protein residues 5 (Ser) to 18 (Ala).  $\beta$  sheet structures run from  
 about 31 (Leu) to 37 (Thr); 42 (Leu) to 44 (Thr); 50 (Ala) to 53  
 (Gly); 63 (Trp) to 66 (Arg); 71 (Cys) to 75 (Cys); and 86 (Cys) to  
 50 89 (Gln). Structure is based upon use of PHD program: accessed by  
<http://www.embl-heidelberg.de/predictprotein/>; or by the DSC  
 program: accessed by <http://bonsai.lif.icnet.uk/bmm/dsc/>. Introns  
 likely about 164-165, and 242-243.

55 GTGTGCCGGA TTTGGTTAGC TGAGCCCACC GAGAGGCGCC TGCAGG ATG AAA GCT  
 55

Met Lys Ala  
 -18

60 CTC TGT CTC CTC CTC CCT GTC CTG GGG CTG TTG GTG TCT AGC AAG  
 103  
 Leu Cys Leu Leu Leu Leu Pro Val Leu Gly Leu Leu Val Ser Ser Lys  
 -15 -10 -5 1



```

ACC CTG TGC TCC ATG GAA GAA GCC ATC AAT GAG AGG ATC CAG GAG GTC
151
Thr Leu Cys Ser Met Glu Glu Ala Ile Asn Glu Arg Ile Gln Glu Val
      5              10              15
5
GCC GGC TCC CTA ATA TTT AGG GCA ATA AGC AGC ATT GGC CTG GAG TGC
199
Ala Gly Ser Leu Ile Phe Arg Ala Ile Ser Ser Ile Gly Leu Glu Cys
      20              25              30
10
CAG AGC GTC ACC TCC AGG GGG GAC CTG GCT ACT TGC CCC CGA GGC TTC
247
Gln Ser Val Thr Ser Arg Gly Asp Leu Ala Thr Cys Pro Arg Gly Phe
      35              40              45
15
GCC GTC ACC GGC TGC ACT TGT GGC TCC GCC TGT GGC TCG TGG GAT GTG
295
Ala Val Thr Gly Cys Thr Cys Gly Ser Ala Cys Gly Ser Trp Asp Val
      50              55              60              65
20
CGC GCC GAG ACC ACA TGT CAC TGC CAG TGC GCG GGC ATG GAC TGG ACC
343
Arg Ala Glu Thr Thr Cys His Cys Gln Cys Ala Gly Met Asp Trp Thr
      70              75              80
25
GGA GCG CGC TGC TGT CGT GTG CAG CCC TGAGGTCGCG CGCAGCGCGT
390
Gly Ala Arg Cys Cys Arg Val Gln Pro
      85              90
30
GCACAGCGCG GCGGAGGCG GCTCCAGGTC CGGAGGGGTT GCGGGGAGC TGGAAATAAA
450
CCT
35 453

```

The CRSPs of this invention are defined in part by  
 their physicochemical and structural properties. The  
 40 biological properties of the mammalian CRSPs described  
 herein, are defined by their amino acid sequence, and  
 mature size. The rodent CRSP molecules, for example,  
 exhibit about 30-46% amino acid identity, depending on  
 whether the signal or mature sequences are compared, and  
 45 somewhere in the range of 55-70+% similarity. One of  
 skill will readily recognize that some sequence  
 variations may be tolerated, e.g., conservative  
 substitutions or positions remote from the helical  
 structures, without altering significantly the biological  
 50 activity of the molecule. The cysteines, being conserved  
 across family members, are probably relatively important  
 structurally. It is likely that most, or all, of them  
 are disulfide linked in important pairings. Note,

however, that the C2 embodiments lack an N proximal cysteine seemingly conserved in the other members described.

In addition, the label may refer, in specific  
5   embodiments, to the nucleic acids which may be isolated  
using PCR amplification from appropriate cells of  
sequences using primers which flank the sequences  
described. Thus, even if minor errors in the given  
10   sequences may have resulted from ambiguities or  
uncertainties in sequencing, the PCR technology will  
allow isolation of natural isolates of the described  
genes. In addition, the nucleotide sequences in the  
regions corresponding to C2 residues 55-62 and 81-88 are  
highly conserved across three mouse embodiments  
15   described. In addition, the nucleotide sequences in the  
regions corresponding to C23 residues 56-63 and 79-86 are  
highly conserved across the known class members. It is  
likely that additional human embodiments will be found  
using appropriate PCR primers encoding such regions.

20   CRSPs are present in specific tissue types, as  
described below. Each correlates with important  
conditions, which are suggestive of important roles in  
immunological conditions. The interaction of the protein  
with a receptor is likely to be important for mediating  
25   various aspects of cellular physiology or development.  
The cellular types which express message encoding CRSPs  
suggest that signals important in cell differentiation  
and development are mediated by them. See, e.g., Gilbert  
(1991) Developmental Biology (3d ed.) Sinauer Associates,  
30   Sunderland, MA; Browder, et al. (1991) Developmental  
Biology (3d ed.) Saunders, Philadelphia, PA.; Russo, et  
al. (1992) Development: The Molecular Genetic Approach  
Springer-Verlag, New York, N.Y.; and Wilkins (1993)  
Genetic Analysis of Animal Development (2d ed.) Wiley-  
35   Liss, New York, N.Y. Moreover, CRSP expression  
correlates with certain specific conditions. See below.

The CRSP producing profile of different cell types  
is elucidated herein. These observations suggest that

the CRSPs represent novel additions to the chemokine/growth factor superfamily.

CRSP protein biochemistry was assessed in some mammalian expression systems. CRSP member C2 was produced as a protein of Mr ~7.5 kDa as evaluated by reducing SDS-PAGE; control transfected supernatants contained no such species. The absence of glycosylation motifs suggests that the natural protein is not glycosylated, and that recombinant protein forms lacking natural glycosylation should share similar biological activity. CRSP member C23 was produced as a protein of Mr ~8 kDa as evaluated by reducing SDS-PAGE; control transfected supernatants contained no such species. The absence of glycosylation motifs suggests that the natural protein is not glycosylated, and that recombinant protein forms lacking natural glycosylation should share similar biological activity.

The cysteine rich nature makes the proteins good substrates for sulfhydryl reagents. They will be useful as a control for cysteine incorporation, and as a sample to test ability to sequence through those residues. The proteins will also find use as a carbon source, in some cases, as labeled.

Since the structure of the proteins are soluble, it is likely that the entire spectrum of inflammatory, infectious, and immunoregulatory states thought to involve other related cytokines or growth factors.

## II. Definitions

The term "binding composition" refers to molecules that bind with specificity to a CRSP, e.g., in an antibody-antigen interaction. However, other compounds, e.g., receptor proteins, may also specifically associate with CRSPs at a higher affinity and/or specificity than other molecules. Typically, the association will be in a natural physiologically relevant protein-protein interaction, either covalent or non-covalent, and may include members of a multiprotein complex, including

carrier compounds or dimerization partners. The molecule may be a polymer, e.g., protein, or chemical reagent. No implication as to whether a CRSP is either a concave or convex surface in a ligand or a receptor of a ligand-receptor interaction is necessarily represented, other than whether the interaction exhibits similar specificity, e.g., specific affinity. A functional analog may be a ligand with structural modifications, e.g., sequence substitutions, often conservative residues, or modified, e.g., derivatized protein polymer, or may be a wholly unrelated molecule, e.g., which has a molecular shape which interacts with the appropriate ligand binding determinants. The ligands may serve as agonists or antagonists of the receptor, see, e.g., Goodman, et al. (eds. 1990) Goodman & Gilman's: The Pharmacological Bases of Therapeutics (8th ed.) Pergamon Press, Tarrytown, N.Y.

The term "binding agent:CRSP complex", as used herein, refers to a complex of a binding agent and a CRSP that is formed by specific binding of the binding agent to the CRSP. Specific binding of the binding agent means that the binding agent has a specific binding site that recognizes a site on the CRSP protein. For example, antibodies raised to a CRSP protein and recognizing an epitope on the CRSP protein are capable of forming a binding agent:CRSP complex by specific binding. Typically, the formation of a binding agent:CRSP complex allows the measurement of CRSP in a mixture of other proteins and biologics. The term "antibody:CRSP complex" refers to an embodiment in which the binding agent is an antibody. The antibody may be monoclonal, polyclonal, or a binding fragment of an antibody, e.g., an Fv, Fab, or F(ab)2 fragment, or even a single chain antibody form. The antibody will often preferably be a polyclonal antibody for cross-reactivity purposes.

"Homologous" nucleic acid sequences, when compared, exhibit significant similarity. The standards for homology in nucleic acids are either measures for homology generally used in the art by sequence comparison

and/or phylogenetic relationship, or based upon hybridization conditions. Hybridization conditions are described in greater detail below.

An "isolated" nucleic acid is a nucleic acid, e.g., an RNA, DNA, or a mixed polymer, which is substantially separated from other biologic components which naturally accompany a native sequence, e.g., proteins and flanking genomic sequences from the originating species. The term embraces a nucleic acid sequence which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs, or analogs biologically synthesized by heterologous systems. A substantially pure molecule includes isolated forms of the molecule. An isolated nucleic acid will usually contain homogeneous nucleic acid molecules, but will, in some embodiments, contain nucleic acids with minor sequence heterogeneity. This heterogeneity is typically found at the polymer ends or portions not critical to a desired biological function or activity.

As used herein, the term "CRSP" shall encompass various specific embodiments, e.g., when used in a protein context, a protein having amino acid sequences, particularly from the cysteine containing portions, shown in SEQ ID NO: 2, 4, 6 or 14, or a significant fragment of such a protein. The invention also embraces a polypeptide which exhibits similar structure to rodent, e.g., mouse CRSP, or primate, e.g. human CRSP, which interacts with CRSP specific binding components. These binding components, e.g., antibodies, typically bind to a CRSP with high affinity, e.g., at least about 100 nM, usually better than about 30 nM, preferably better than about 10 nM, and more preferably at better than about 3 nM.

The term "polypeptide" or "protein" as used herein includes a significant fragment or segment of the conserved cysteine containing portions of a CRSP, and encompasses a stretch of amino acid residues of at least about 8 amino acids, generally at least 10 amino acids,

more generally at least 12 amino acids, often at least 14 amino acids, more often at least 16 amino acids, typically at least 18 amino acids, more typically at least 20 amino acids, usually at least 22 amino acids, more usually at least 24 amino acids, preferably at least 26 amino acids, more preferably at least 28 amino acids, and, in particularly preferred embodiments, at least about 30 or more amino acids, e.g., 35, 40, 45, 50, 60, 70, 80, etc. Particularly important segments include those which span over regions between, e.g., the residues corresponding to C2 residues 32-85, or the residues corresponding to C23 residues 33-86 or significant portions thereof. Thus, fragments corresponding to ends beginning at 35, 36, 37, etc., and ending, e.g., at positions 80, 79, 78, 77, etc., will be interesting. Antigenic fragments will be utilized, e.g., in generating antibody reagents.

Typical embodiments will exhibit a plurality of distinct, e.g., nonoverlapping, segments of the specified length. Typically, the plurality will be at least two, more usually at least three, and preferably 5, 7, or even more. While the length minima are provided, longer lengths, of various sizes, may be appropriate, e.g., one of length 7, and two of length 12.

A "recombinant" nucleic acid is defined either by its method of production or its structure. In reference to its method of production, e.g., a product made by a process, the process is use of recombinant nucleic acid techniques, e.g., involving human intervention in the nucleotide sequence, typically selection and/or production. Alternatively, it can be a nucleic acid made by generating a sequence comprising fusion of two fragments which are not naturally contiguous to each other, but is meant to exclude products of nature, e.g., naturally occurring mutants. Thus, for example, products made by transforming cells with any non-naturally occurring vector is encompassed, as are nucleic acids comprising sequence derived using any synthetic oligonucleotide process. Such is often done to replace a

codon with a redundant codon encoding the same or a similar amino acid, while typically introducing or removing a sequence recognition site. Preferred embodiments include, e.g., 1-fold, 2-fold, 3-fold, 5-  
5 fold, 7-fold, etc., preferably conservative substitutions at the nucleotide or amino acid levels. Preferably the substitutions will be away from the conserved cysteines, and often will be in the regions away from the helical structural domains. Such variants may be useful to  
10 produce specific antibodies, and often will share many or all biological properties.

Alternatively, recombinant manipulation is performed to join together nucleic acid segments of desired functions to generate a single genetic entity comprising  
15 a desired combination of functions not found in the commonly available natural forms. Restriction enzyme recognition sites are often the target of such artificial manipulations, but other site specific targets, e.g., promoters, DNA replication sites, regulation sequences,  
20 control sequences, or other useful features may be incorporated by design. A similar concept is intended for a recombinant, e.g., fusion, polypeptide. Specifically included are synthetic nucleic acids which, by genetic code redundancy, encode polypeptides similar  
25 to fragments of these antigens, and fusions of sequences from various different species variants.

"Solubility" is reflected by sedimentation measured in Svedberg units, which are a measure of the sedimentation velocity of a molecule under particular  
30 conditions. The determination of the sedimentation velocity was classically performed in an analytical ultracentrifuge, but is typically now performed in a standard ultracentrifuge. See, Freifelder (1982) Physical Biochemistry (2d ed.) W.H. Freeman & Co., San  
35 Francisco, CA; and Cantor and Schimmel (1980) Biophysical Chemistry parts 1-3, W.H. Freeman & Co., San Francisco, CA. As a crude determination, a sample containing a putatively soluble polypeptide is spun in a standard full sized ultracentrifuge at about 50K rpm for about 10

minutes, and soluble molecules will remain in the supernatant. A soluble particle or polypeptide will typically be less than about 30S, more typically less than about 15S, usually less than about 10S, more usually less than about 6S, and, in particular embodiments, preferably less than about 4S, and more preferably less than about 3S. Solubility of a polypeptide or fragment depends upon the environment and the polypeptide. Many parameters affect polypeptide solubility, including temperature, electrolyte environment, size and molecular characteristics of the polypeptide, and nature of the solvent. Typically, the temperature at which the polypeptide is used ranges from about 4° C to about 65° C. Usually the temperature at use is greater than about 18° C and more usually greater than about 22° C. For diagnostic purposes, the temperature will usually be about room temperature or warmer, but less than the denaturation temperature of components in the assay. For therapeutic purposes, the temperature will usually be body temperature, typically about 37° C for humans, though under certain situations the temperature may be raised or lowered in situ or in vitro.

The size and structure of the polypeptide should generally be in a substantially stable state, and usually not in a denatured state. The polypeptide may be associated with other polypeptides in a quaternary structure, e.g., to confer solubility, or associated with lipids or detergents in a manner which approximates natural lipid bilayer interactions. In other contexts, the protein may be denatured, e.g., in Western protein blot analysis, or to minimize certain tertiary conformation features of sequences.

The solvent will usually be a biologically compatible buffer, of a type used for preservation of biological activities, and will usually approximate a physiological solvent. Usually the solvent will have a neutral pH, typically between about 5 and 10, and preferably about 7.5. On some occasions, a detergent will be added, typically a mild non-denaturing one, e.g.,



CHS (cholesteryl hemisuccinate) or CHAPS (3-[3-cholamidopropyl]dimethylammonio]-1-propane sulfonate), or a low enough concentration as to avoid significant disruption of structural or physiological properties of the protein. Sterile compositions are often useful, e.g., in a tissue culture assay context.

"Substantially pure" in a protein context typically means that the protein is isolated from other contaminating proteins, nucleic acids, and other biologicals derived from the original, e.g., natural, source organism. Purity, or "isolation" may be assayed by standard methods, and will ordinarily be at least about 50% pure, more ordinarily at least about 60% pure, generally at least about 70% pure, more generally at least about 80% pure, often at least about 85% pure, more often at least about 90% pure, preferably at least about 95% pure, more preferably at least about 98% pure, and in most preferred embodiments, at least 99% pure. The measure will typically be by mass, but may be molar. Similar concepts apply, e.g., to antibodies or nucleic acids.

"Substantial similarity" in the nucleic acid sequence comparison context means either that the segments, or their complementary strands, when compared, are identical when optimally aligned, with appropriate nucleotide insertions or deletions, in at least about 50% of the nucleotides, generally at least 56%, more generally at least 59%, ordinarily at least 62%, more ordinarily at least 65%, often at least 68%, more often at least 71%, typically at least 74%, more typically at least 77%, usually at least 80%, more usually at least about 85%, preferably at least about 90%, more preferably at least about 95 to 98% or more, and in particular embodiments, as high at about 99% or more of the nucleotides. Alternatively, substantial similarity exists when the segments will hybridize under selective hybridization conditions, to a strand, or its complement, typically using a sequence derived from SEQ ID NO: 1, 3, or 5. Typically, selective hybridization will occur when

there is at least about 55% similarity over a stretch of at least about 30 nucleotides, preferably at least about 65% over a stretch of at least about 25 nucleotides, more preferably at least about 75%, and most preferably at least about 90% over about 20 nucleotides. See Kanehisa (1984) Nuc. Acids Res. 12:203-213. The length of similarity comparison, as described, may be over longer stretches, and in certain embodiments will be over a stretch of at least about 17 nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 40 nucleotides, preferably at least about 50 nucleotides, and more preferably at least about 75 to 100 or more nucleotides, e.g., 150, 200, etc.

"Stringent conditions", in referring to homology or substantial similarity in the hybridization context, will be stringent combined conditions of salt, temperature, organic solvents, and other parameters, typically those controlled in hybridization reactions. The combination of parameters is more important than the measure of any single parameter. See, e.g., Wetmur and Davidson (1968) J. Mol. Biol. 31:349-370. A nucleic acid probe which binds to a target nucleic acid under stringent conditions is specific, e.g., for hybridizing to said target nucleic acid. Such a probe is typically more than 11 nucleotides in length, and is sufficiently identical or complementary to a target nucleic acid over the region specified by the sequence of the probe to bind the target under stringent hybridization conditions. Hybridization under stringent conditions should give a background of at least 2-fold over background, preferably at least 3-5 or more. PCR primers are often prime examples of embodiments where specificity of relatively short sequence segments will be used.

CRSPs from other mammalian species can be cloned and isolated by cross-species hybridization of closely related species. See, e.g., below. Alternatively, other related proteins may be isolated by protein purification

or antigenic methods using cross-reacting antibody reagents. It also appears that the genes may be genomically clustered within a species, so other members of the family may be isolated by positional cloning techniques upon identification of one embodiment. Similarity may be relatively low between distantly related species, and thus hybridization of relatively closely related species is advisable. Alternatively, preparation of an antibody preparation which exhibits less species specificity may be useful in expression cloning or protein isolation approaches.

The phrase "specifically binds to an antibody" or "specifically immunoreactive with", when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biological components. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not significantly bind other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised to a rodent CRSP immunogen with the amino acid sequence depicted in SEQ ID NO: 2, 4, or 6 can be selected to obtain antibodies specifically immunoreactive with CRSP proteins and not with other proteins. These antibodies will recognize proteins highly similar to the homologous mouse CRSP protein or proteins, as selected.

30

### III. Nucleic Acids

Rodent CRSPs and the C23 human CRSP are exemplary of a larger class of structurally and functionally related proteins. These soluble proteins will likely serve to transmit signals between different cell types. The preferred embodiments, as disclosed, will be useful in standard procedures to isolate genes from different individuals or other species, e.g., warm blooded animals, such as birds and mammals. Cross hybridization or other

techniques will allow isolation of related genes encoding proteins from individuals, strains, or species. In fact, Applicants possess specific data where mouse to rat cross species hybridization has been strongly indicated. A  
5 number of different approaches are available to successfully isolate a suitable nucleic acid clone based upon the information provided herein.

While the overall coding sequence identity among the mouse family members is in the 40% range overall,  
10 specific portions exhibit over 80%. Southern blot hybridization studies using specific portions might qualitatively determine the presence of homologous genes in human, monkey, rat, dog, cow, and rabbit genomes under specific hybridization conditions. PCR techniques may be  
15 useful using, e.g., the conserved sequence regions, preferably at the nucleotide level, but perhaps also at the protein level.

Complementary sequences will also be used as probes or primers. Based upon identification of the likely  
20 amino terminus, other peptides should be particularly useful, e.g., coupled with anchored vector or poly-A complementary PCR techniques or with complementary DNA of other peptides.

Techniques for nucleic acid manipulation of genes  
25 encoding CRSP proteins, such as subcloning nucleic acid sequences encoding polypeptides into expression vectors, labeling probes, DNA hybridization, and the like are described generally in Sambrook, et al. (1989) Molecular Cloning: A Laboratory Manual (2nd ed.) Vol. 1-3, Cold  
30 Spring Harbor Laboratory, Cold Spring Harbor Press, NY, which is incorporated herein by reference. This manual is hereinafter referred to as "Sambrook, et al."

There are various methods of isolating DNA sequences encoding CRSPs. For example, DNA is isolated from a  
35 genomic or cDNA library using labeled oligonucleotide probes having sequences identical or complementary to the sequences disclosed herein. Full-length probes may be used, or oligonucleotide probes may be generated by comparison of the sequences disclosed. Such probes can

be used directly in hybridization assays to isolate DNA encoding CRSPs, or probes can be designed for use in amplification techniques such as PCR, for the isolation of DNA encoding CRSPs. Certain genomic searches of  
5 available sequence databases, public or private, will also be useful to identify additional members.

To prepare a cDNA library, mRNA is isolated from cells which express a CRSP protein. cDNA is prepared from the mRNA and ligated into a recombinant vector. The  
10 vector is transfected into a recombinant host for propagation, screening, and cloning. Methods for making and screening cDNA libraries are well known. See Gubler and Hoffman (1983) Gene 25:263-269 and Sambrook, et al.

For a genomic library, e.g., the DNA can be  
15 extracted from tissue and either mechanically sheared or enzymatically digested to yield fragments of about 12-20 kb. The fragments are then separated by gradient centrifugation and cloned in bacteriophage lambda vectors. These vectors and phage are packaged in vitro,  
20 as described in Sambrook, et al. Recombinant phage are analyzed by plaque hybridization as described in Benton and Davis (1977) Science 196:180-182. Colony hybridization is carried out as generally described in e.g., Grunstein, et al. (1975) Proc. Natl. Acad. Sci.  
25 USA. 72:3961-3965. Modifications may be incorporated.

DNA encoding a CRSP can be identified in either cDNA or genomic libraries by its ability to hybridize with the nucleic acid probes described herein, e.g., in colony or plaque hybridization assays. The corresponding DNA  
30 regions are isolated by standard methods familiar to those of skill in the art. See, e.g., Sambrook, et al.

Various methods of amplifying target sequences, such as the polymerase chain reaction, can also be used to prepare DNA encoding CRSPs. Polymerase chain reaction  
35 (PCR) technology is used to amplify such nucleic acid sequences directly from mRNA, from cDNA, and from genomic libraries or cDNA libraries. The isolated sequences encoding CRSP proteins may also be used as templates for

PCR amplification. The sequences provided teach many appropriate primers, and pairs.

Typically, in PCR techniques, oligonucleotide primers complementary to two 5' regions in the DNA region to be amplified are synthesized. The polymerase chain reaction is then carried out using the two primers. See Innis, et al. (eds. 1990) PCR Protocols: A Guide to Methods and Applications Academic Press, San Diego, CA. Primers can be selected to amplify the entire regions encoding a full-length CRSP protein or to amplify smaller DNA segments as desired. Once such regions are PCR-amplified, they can be sequenced and oligonucleotide probes can be prepared from sequence obtained using standard techniques. These probes can then be used to isolate DNA's encoding CRSP proteins.

Oligonucleotides for use as probes are usually chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage and Carruthers (1983) Tetrahedron Lett. 22:1859-1862, or using an automated synthesizer, as described in Needham-VanDevanter, et al. (1984) Nucleic Acids Res. 12:6159-6168. Purification of oligonucleotides is performed, e.g., by native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson and Regnier (1983) J. Chrom. 255:137-149. The sequence of the synthetic oligonucleotide can be verified using, e.g., the chemical degradation method of Maxam, A.M. and Gilbert, W. in Grossman and Moldave (eds. 1980) Methods in Enzymology 65:499-560 Academic Press, New York.

Isolated nucleic acids encoding rodent CRSPs have been identified. The nucleotide sequence and corresponding open reading frame of various embodiments are provided in SEQ ID NO: 1-12.

An isolated nucleic acid encoding a human CRSP has also been identified. The nucleotide sequence and corresponding open reading frame of this embodiment is provided in SEQ ID NO: 13.

Notably, the different genes of the family exhibit great conservation of spacing of the cysteine residues. In particular, the spacing is CX11CX8CXCX3CX10CXCX9CC. See Table 5.

5       Based upon the structural modeling and insights in the binding regions of the collective group, it is predicted that residues in the mature protein, lacking a signal of about 20-23 residues (see, e.g., von Heijne (1986) Nucl. Acids Res. 14:4683-4691; and Nielsen, et al. 10 (1997) Protein Eng. 10:1-9), exhibit structural features as described in Table 1. The soluble protein has been predicted to possess one helical structure, and six beta strands. The structure suggests that the sheets form a plane, and that the underside of the plane is 15 probably the receptor binding surface. Thus, residue substitutions at positions on the upper surface, or in the helical region away from the sheet surface will not affect biological activity.

      Fragments of at least about 8-10 residues in the 20 cysteine rich region would be especially interesting peptides, e.g., starting at residue positions of the mature protein 1, 2, 3, etc. However, the cysteine rich region begins at 33-86. Those fragments would typically end at the ends of the helical or sheet strands should be 25 important. Other interesting peptides of various lengths would include ones which begin or end in other positions of the protein, e.g., at residues 87, 86, etc., with lengths ranging, e.g., from about 8 to 20, 25, 30, 35, 40, etc.

30       This invention provides isolated DNA or fragments to encode a CRSP protein. In addition, this invention provides isolated or recombinant DNA which encodes a protein or polypeptide which is capable of hybridizing under appropriate conditions, e.g., high stringency, with 35 the DNA sequences described herein. Said biologically active protein or polypeptide can be an intact ligand, or fragment, and have an amino acid sequence as disclosed in SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14. Preferred embodiments will be full length natural sequences, from

isolates, e.g., about 7-8K daltons in size when unglycosylated, or fragments of at least about 1,000 daltons, more preferably at least about 3,000 daltons. In glycosylated forms, which appear unnatural, the protein may be larger. Further, this invention contemplates the use of isolated or recombinant DNA, or fragments thereof, which encode proteins which are homologous to a CRSP or which were isolated using cDNA encoding a CRSP as a probe. The isolated DNA can have the respective regulatory sequences in the 5' and 3' flanks, e.g., promoters, enhancers, poly-A addition signals, and others.

#### IV. Making CRSPs

DNAs which encode a CRSP or fragments thereof can be obtained by chemical synthesis, screening cDNA libraries, or by screening genomic libraries prepared from a wide variety of cell lines or tissue samples.

These DNAs can be expressed in a wide variety of host cells for the synthesis of a full-length protein or fragments which can in turn, e.g., be used to generate polyclonal or monoclonal antibodies; for binding studies; for construction and expression of modified molecules; and for structure/function studies. Each CRSP or its fragments can be expressed in host cells that are transformed or transfected with appropriate expression vectors. These molecules can be substantially purified to be free of protein or cellular contaminants, other than those derived from the recombinant host, and therefore are particularly useful in pharmaceutical compositions when combined with a pharmaceutically acceptable carrier and/or diluent. The antigen, e.g., CRSP, or portions thereof, may be expressed as fusions with other proteins or possessing an epitope tag. Forms with a carboxy terminal FLAG epitope tag have been produced.

Expression vectors are typically self-replicating DNA or RNA constructs containing the desired antigen gene or its fragments, usually operably linked to appropriate



genetic control elements that are recognized in a suitable host cell. The specific type of control elements necessary to effect expression will depend upon the eventual host cell used. Generally, the genetic control elements can include a prokaryotic promoter system or a eukaryotic promoter expression control system, and typically include a transcriptional promoter, an optional operator to control the onset of transcription, transcription enhancers to elevate the level of mRNA expression, a sequence that encodes a suitable ribosome binding site, and sequences that terminate transcription and translation. Expression vectors also usually contain an origin of replication that allows the vector to replicate independently from the host cell.

The vectors of this invention contain DNAs which encode a CRSP, or a fragment thereof, typically encoding, e.g., a biologically active polypeptide, or protein. The DNA can be under the control of a viral promoter and can encode a selection marker. This invention further contemplates use of such expression vectors which are capable of expressing eukaryotic cDNA coding for a CRSP in a prokaryotic or eukaryotic host, where the vector is compatible with the host and where the eukaryotic cDNA coding for the protein is inserted into the vector such that growth of the host containing the vector expresses the cDNA in question. Usually, expression vectors are designed for stable replication in their host cells or for amplification to greatly increase the total number of copies of the desirable gene per cell. It is not always necessary to require that an expression vector replicate in a host cell, e.g., it is possible to effect transient expression of the protein or its fragments in various hosts using vectors that do not contain a replication origin that is recognized by the host cell. It is also possible to use vectors that cause integration of a CRSP gene or its fragments into the host DNA by recombination, or to integrate a promoter which controls expression of an endogenous gene.

Vectors, as used herein, contemplate plasmids, viruses, bacteriophage, integratable DNA fragments, and other vehicles which enable the integration of DNA fragments into the genome of the host. Expression  
5 vectors are specialized vectors which contain genetic control elements that effect expression of operably linked genes. Plasmids are the most commonly used form of vector, but many other forms of vectors which serve an equivalent function are suitable for use herein. See,  
10 e.g., Pouwels, et al. (1985 and Supplements) Cloning Vectors: A Laboratory Manual Elsevier, N.Y.; and Rodriguez, et al. (eds. 1988) Vectors: A Survey of Molecular Cloning Vectors and Their Uses Buttersworth, Boston, MA.

15 Suitable host cells include prokaryotes, lower eukaryotes, and higher eukaryotes. Prokaryotes include both gram negative and gram positive organisms, e.g., E. coli and B. subtilis. Lower eukaryotes include yeasts, e.g., S. cerevisiae and Pichia, and species of the genus  
20 Dictyostelium. Higher eukaryotes include established tissue culture cell lines from animal cells, both of non-mammalian origin, e.g., insect cells, and birds, and of mammalian origin, e.g., human, primates, and rodents.

Prokaryotic host-vector systems include a wide  
25 variety of vectors for many different species. As used herein, E. coli and its vectors will be used generically to include equivalent vectors used in other prokaryotes. A representative vector for amplifying DNA is pBR322 or its derivatives. Vectors that can be used to express  
30 CRSPs or CRSP fragments include, but are not limited to, such vectors as those containing the lac promoter (pUC-series); trp promoter (pBR322-trp); Ipp promoter (the pIN-series); lambda-pP or pR promoters (pOTS); or hybrid promoters such as ptac (pDR540). See Brosius, et al.  
35 (1988) "Expression Vectors Employing Lambda-, trp-, lac-, and Ipp-derived Promoters", in Rodriguez and Denhardt (eds.) Vectors: A Survey of Molecular Cloning Vectors and Their Uses 10:205-236 Buttersworth, Boston, MA.

Lower eukaryotes, e.g., yeasts and Dictyostelium, may be transformed with CRSP sequence containing vectors. For purposes of this invention, the most common lower eukaryotic host is the baker's yeast, *Saccharomyces cerevisiae*. It will be used generically to represent lower eukaryotes although a number of other strains and species are also available. Yeast vectors typically consist of a replication origin (unless of the integrating type), a selection gene, a promoter, DNA encoding the desired protein or its fragments, and sequences for translation termination, polyadenylation, and transcription termination. Suitable expression vectors for yeast include such constitutive promoters as 3-phosphoglycerate kinase and various other glycolytic enzyme gene promoters or such inducible promoters as the alcohol dehydrogenase 2 promoter or metallothionine promoter. Suitable vectors include derivatives of the following types: self-replicating low copy number (such as the YRp-series), self-replicating high copy number (such as the YEp-series); integrating types (such as the YIp-series), or mini-chromosomes (such as the YCp-series).

Higher eukaryotic tissue culture cells are typically the preferred host cells for expression of the functionally active CRSP protein. In principle, many higher eukaryotic tissue culture cell lines may be used, e.g., insect baculovirus expression systems, whether from an invertebrate or vertebrate source. However, mammalian cells are preferred to achieve proper processing, both cotranslationally and posttranslationally. Transformation or transfection and propagation of such cells is routine. Useful cell lines include HeLa cells, Chinese hamster ovary (CHO) cell lines, baby rat kidney (BRK) cell lines, insect cell lines, bird cell lines, and monkey (COS) cell lines. Expression vectors for such cell lines usually include an origin of replication, a promoter, a translation initiation site, RNA splice sites (e.g., if genomic DNA is used), a polyadenylation site, and a transcription termination site. These vectors also

may contain a selection gene or amplification gene.

Suitable expression vectors may be plasmids, viruses, or retroviruses carrying promoters derived, e.g., from such sources as from adenovirus, SV40, parvoviruses, vaccinia virus, or cytomegalovirus. Representative examples of suitable expression vectors include pCDNA1; pCD, see Okayama, et al. (1985) Mol. Cell Biol. 5:1136-1142; pMC1neo Poly-A, see Thomas, et al. (1987) Cell 51:503-512; and a baculovirus vector such as pAC 373 or pAC 610.

10 It is likely that CRSPs need not be glycosylated to elicit biological responses. However, it will occasionally be desirable to express a CRSP polypeptide in a system which provides a specific or defined glycosylation pattern. In this case, the usual pattern will be that provided naturally by the expression system. However, the pattern will be modifiable by exposing the polypeptide, e.g., in unglycosylated form, to appropriate glycosylating proteins introduced into a heterologous expression system. For example, the CRSP gene may be co-transformed with one or more genes encoding mammalian or other glycosylating enzymes. It is further understood that over glycosylation may be detrimental to CRSP biological activity, and that one of skill may perform routine testing to optimize the degree of glycosylation which confers optimal biological activity.

25 A CRSP, or a fragment thereof, may be engineered to be phosphatidyl inositol (PI) linked to a cell membrane, but can be removed from membranes by treatment with a phosphatidyl inositol cleaving enzyme, e.g., phosphatidyl inositol phospholipase-C. This releases the antigen in a biologically active form, and allows purification by standard procedures of protein chemistry. See, e.g., Low (1989) Biochem. Biophys. Acta 988:427-454; Tse, et al. (1985) Science 230:1003-1008; and Brunner, et al. (1991) J. Cell Biol. 114:1275-1283.

Now that CRSPs have been characterized, fragments or derivatives thereof can be prepared by conventional processes for synthesizing peptides. These include processes such as are described in Stewart and Young

(1984) Solid Phase Peptide Synthesis Pierce Chemical Co., Rockford, IL; Bodanszky and Bodanszky (1984) The Practice of Peptide Synthesis Springer-Verlag, New York, NY; and Bodanszky (1984) The Principles of Peptide Synthesis  
5 Springer-Verlag, New York, NY. For example, an azide process, an acid chloride process, an acid anhydride process, a mixed anhydride process, an active ester process (for example, p-nitrophenyl ester, N-hydroxysuccinimide ester, or cyanomethyl ester), a  
10 carbodiimidazole process, an oxidative-reductive process, or a dicyclohexylcarbodiimide (DCCD)/additive process can be used. Solid phase and solution phase syntheses are both applicable to the foregoing processes.

The prepared protein and fragments thereof can be  
15 isolated and purified from the reaction mixture by means of peptide separation, for example, by extraction, precipitation, electrophoresis and various forms of chromatography, and the like. The CRSPs of this invention can be obtained in varying degrees of purity  
20 depending upon its desired use. Purification can be accomplished by use of known protein purification techniques or by the use of the antibodies or binding partners herein described, e.g., in immunoabsorbant affinity chromatography. This immunoabsorbant affinity  
25 chromatography is carried out by first linking the antibodies to a solid support and then contacting the linked antibodies with solubilized lysates of appropriate source cells, lysates of other cells expressing the ligand, or lysates or supernatants of cells producing the  
30 CRSPs as a result of recombinant DNA techniques, see below.

Multiple cell lines may be screened for one which expresses a CRSP at a high level compared with other cells. Various cell lines, e.g., a mouse thymic stromal  
35 cell line TA4, is screened and selected for its favorable handling properties. Natural CRSPs can be isolated from natural sources, or by expression from a transformed cell using an appropriate expression vector. Purification of the expressed protein is achieved by standard procedures,

or may be combined with engineered means for effective purification at high efficiency from cell lysates or supernatants. Epitope or other tags, e.g., FLAG or His<sub>6</sub> segments, can be used for such purification features.

5

## V. Antibodies

Antibodies can be raised to various CRSPs, including individual, polymorphic, allelic, strain, or species variants, and fragments thereof, both in their naturally occurring (full-length) forms and in their recombinant forms. Additionally, antibodies can be raised to CRSPs in either their active forms or in their inactive forms. Anti-idiotypic antibodies may also be used.

### A. Antibody Production

15 A number of immunogens may be used to produce antibodies specifically reactive with CRSP proteins. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring protein may also be used either in  
20 pure or impure form. Synthetic peptides, made using the mouse CRSP protein sequences described herein, may also be used as an immunogen for the production of antibodies to CRSPs. Recombinant protein can be expressed in eukaryotic or prokaryotic cells as described herein, and  
25 purified as described. Naturally folded or denatured material can be used, as appropriate, for producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the protein.

30 Methods of producing polyclonal antibodies are known to those of skill in the art. Typically, an immunogen, preferably a purified protein, is mixed with an adjuvant and animals are immunized with the mixture. The animal's immune response to the immunogen preparation is monitored  
35 by taking test bleeds and determining the titer of reactivity to the CRSP protein of interest. When appropriately high titers of antibody to the immunogen are obtained, usually after repeated immunizations, blood is collected from the animal and antisera are prepared.

Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired. See, e.g., Harlow and Lane; or Coligan.

- Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Typically, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (see, Kohler and Milstein (1976) Eur. J. Immunol. 6:511-519, incorporated herein by reference).
- Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host.
- Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according, e.g., to the general protocol outlined by Huse, et al. (1989) Science 246:1275-1281.

- Antibodies, including binding fragments and single chain versions, against predetermined fragments of CRSPs can be raised by immunization of animals with conjugates of the fragments with carrier proteins as described above. Monoclonal antibodies are prepared from cells secreting the desired antibody. These antibodies can be screened for binding to normal or defective CRSPs, or screened for agonistic or antagonistic activity, e.g., mediated through a receptor. These monoclonal antibodies will usually bind with at least a  $K_D$  of about 1 mM, more usually at least about 300  $\mu$ M, typically at least about 10  $\mu$ M, more typically at least about 30  $\mu$ M, preferably at least about 10  $\mu$ M, and more preferably at least about 3  $\mu$ M or better.

In some instances, it is desirable to prepare monoclonal antibodies from various mammalian hosts, such

as mice, rodents, primates, humans, etc. Description of techniques for preparing such monoclonal antibodies may be found in, e.g., Stites, et al. (eds.) Basic and Clinical Immunology (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane (1988) Antibodies: A Laboratory Manual CSH Press; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.) Academic Press, New York, NY; and particularly in Kohler and Milstein (1975) Nature 256:495-497, which discusses one method of generating monoclonal antibodies. Summarized briefly, this method involves injecting an animal with an immunogen. The animal is then sacrificed and cells taken from its spleen, which are then fused with myeloma cells. The result is a hybrid cell or "hybridoma" that is capable of reproducing in vitro. The population of hybridomas is then screened to isolate individual clones, each of which secrete a single antibody species to the immunogen. In this manner, the individual antibody species obtained are the products of immortalized and cloned single B cells from the immune animal generated in response to a specific site recognized on the immunogenic substance.

Other suitable techniques involve selection of libraries of antibodies in phage or similar vectors. See, e.g., Huse, et al. (1989) "Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire in Phage Lambda," Science 246:1275-1281; and Ward, et al. (1989) Nature 341:544-546. The polypeptides and antibodies of the present invention may be used with or without modification, including chimeric or humanized antibodies. Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents,



teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins may be produced. See, Cabilly, U.S. Patent No. 4,816,567; and Queen, et al. (1989) Proc. Nat'l Acad. Sci. USA 86:10029-10033.

The antibodies of this invention are useful for affinity chromatography in isolating CRSP protein. Columns can be prepared where the antibodies are linked to a solid support, e.g., particles, such as agarose, SEPHADEX, or the like, where a cell lysate or supernatant may be passed through the column, the column washed, followed by increasing concentrations of a mild denaturant, whereby purified CRSP protein will be released.

The antibodies may also be used to screen expression libraries for particular expression products. Usually the antibodies used in such a procedure will be labeled with a moiety allowing easy detection of presence of antigen by antibody binding.

Antibodies to CRSPs may be used for the identification of cell populations expressing CRSPs. By assaying the expression products of cells expressing CRSPs it is possible to diagnose disease, e.g., immune-compromised conditions.

Antibodies raised against each CRSP will also be useful to raise anti-idiotypic antibodies. These will be useful in detecting or diagnosing various immunological conditions related to expression of the respective antigens.

#### B. Immunoassays

A particular protein can be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures in general, see Stites and Terr (eds. 1991) Basic and Clinical Immunology (7th ed.). Moreover, the immunoassays of the present invention can be performed in many configurations, which are reviewed extensively in Maggio (ed. 1980) Enzyme Immunoassay CRC Press, Boca Raton, Florida; Tijan (1985) "Practice and

Theory of Enzyme Immunoassays," Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers B.V., Amsterdam; and Harlow and Lane Antibodies, A Laboratory Manual, supra, each of which is  
5 incorporated herein by reference. See also Chan (ed. 1987) Immunoassay: A Practical Guide Academic Press, Orlando, FL; Price and Newman (eds. 1991) Principles and Practice of Immunoassays Stockton Press, NY; and Ngo (ed. 1988) Non-isotopic Immunoassays Plenum Press, NY.

10 Immunoassays for measurement of CRSP proteins can be performed by a variety of methods known to those skilled in the art. In brief, immunoassays to measure the protein can be either competitive or noncompetitive binding assays. In competitive binding assays, the  
15 sample to be analyzed competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is an antibody specifically reactive with CRSP proteins produced as described above. The concentration of  
20 labeled analyte bound to the capture agent is inversely proportional to the amount of free analyte present in the sample.

In a competitive binding immunoassay, the CRSP protein present in the sample competes with labeled  
25 protein for binding to a specific binding agent, for example, an antibody specifically reactive with the CRSP protein. The binding agent may be bound to a solid surface to effect separation of bound labeled protein from the unbound labeled protein. Alternately, the  
30 competitive binding assay may be conducted in liquid phase and a variety of techniques known in the art may be used to separate the bound labeled protein from the unbound labeled protein. Following separation, the amount of bound labeled protein is determined. The  
35 amount of protein present in the sample is inversely proportional to the amount of labeled protein binding.

Alternatively, a homogeneous immunoassay may be performed in which a separation step is not needed. In these immunoassays, the label on the protein is altered

by the binding of the protein to its specific binding agent. This alteration in the labeled protein results in a decrease or increase in the signal emitted by label, so that measurement of the label at the end of the  
5 immunoassay allows for detection or quantitation of the protein."

Quantitation of CRSP proteins may also be performed using many of a variety of noncompetitive immunoassay methods. For example, a two-site, solid phase sandwich  
10 immunoassay may be used. In this type of assay, a binding agent for the protein, for example an antibody, is attached to a solid support. A second protein binding agent, which may also be an antibody, and which binds the protein at a different site, is labeled. After binding  
15 at both sites on the protein has occurred, the unbound labeled binding agent is removed and the amount of labeled binding agent bound to the solid phase is measured. The amount of labeled binding agent bound is directly proportional to the amount of protein in the  
20 sample.

Western blot analysis can be used to determine the presence of CRSP proteins in a sample. Electrophoresis is carried out, for example, on a tissue sample suspected of containing the protein. Following electrophoresis to  
25 separate the proteins, and transfer of the proteins to a suitable solid support, e.g., a nitrocellulose filter, the solid support is incubated with an antibody reactive with the protein. This antibody may be labeled, or alternatively may be detected by subsequent incubation  
30 with a second labeled antibody that binds the primary antibody.

The immunoassay formats described above employ labeled assay components. The label may be coupled directly or indirectly to the desired component of the  
35 assay according to methods well known in the art. A wide variety of labels and methods may be used.

Traditionally, a radioactive label incorporating  $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^{32}\text{P}$  was used. Non-radioactive labels include ligands which bind to labeled antibodies,

fluorophores, chemiluminescent agents, enzymes, and antibodies which can serve as specific binding pair members for a labeled ligand. The choice of label depends on sensitivity required, ease of conjugation with the compound, stability requirements, and available instrumentation. For a review of various labeling or signal producing systems which may be used, see U.S. Patent No. 4,391,904, which is incorporated herein by reference.

Antibodies reactive with a particular protein can also be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures applicable to the measurement of antibodies by immunoassay techniques, see Stites and Terr (eds.) Basic and Clinical Immunology (7th ed.) supra; Maggio (ed.) Enzyme Immunoassay, supra; and Harlow and Lane Antibodies, A Laboratory Manual, supra.

In brief, immunoassays to measure antisera reactive with CRSP proteins can be either competitive or noncompetitive binding assays. In competitive binding assays, the sample analyte competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is a purified recombinant CRSP protein produced as described above. Other sources of CRSP proteins, including isolated or partially purified naturally occurring protein, may also be used. Noncompetitive assays include sandwich assays, in which the sample analyte is bound between two analyte-specific binding reagents. One of the binding agents is used as a capture agent and is bound to a solid surface. The second binding agent is labeled and is used to measure or detect the resultant complex by visual or instrument means. A number of combinations of capture agent and labeled binding agent can be used. A variety of different immunoassay formats, separation techniques, and labels can be also be used similar to those described above for the measurement of CRSP proteins.

## VI. Purified CRSPs

Specific embodiments of mouse CRSP amino acid sequences are provided in SEQ ID NO: 2, 4, 6, or 8. A rat C19 counterpart is described in SEQ ID NO: 10, and human related C10 and C23 proteins are described in SEQ ID NO: 12 and 14.

Purified protein or defined peptides are useful for generating antibodies by standard methods, as described above. Synthetic peptides or purified protein can be presented to an immune system to generate polyclonal and monoclonal antibodies. See, e.g., Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; and Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press, NY, which are incorporated herein by reference. Alternatively, a CRSP receptor can be useful as a specific binding reagent, and advantage can be taken of its specificity of binding, for, e.g., purification of a CRSP ligand.

The specific binding composition can be used for screening an expression library made from a cell line which expresses a CRSP. Many methods for screening are available, e.g., standard staining of surface expressed ligand, or by panning. Screening of intracellular expression can also be performed by various staining or immunofluorescence procedures. The binding compositions could be used to affinity purify or sort out cells expressing the ligand.

The peptide segments, along with comparison to homologous genes, can also be used to produce appropriate oligonucleotides to screen a library. The genetic code can be used to select appropriate oligonucleotides useful as probes for screening. In combination with polymerase chain reaction (PCR) techniques, synthetic oligonucleotides will be useful in selecting desired clones from a library, including natural allelic and polymorphic variants.

The peptide sequences allow preparation of peptides to generate antibodies to recognize such segments, and allow preparation of oligonucleotides which encode such

sequences. The sequence also allows for synthetic preparation, e.g., see Dawson, et al. (1994) Science 266:776-779. Since CRSPs appear to be secreted proteins, each gene will normally possess an N-terminal signal sequence, which is removed upon processing and secretion, and the putative cleavage site is experimentally determined or predicted as shown in Table 1 through 6, though the cleavage positions in a given host cell may be slightly in either direction.

#### VII. Physical Variants

This invention also encompasses proteins or peptides having substantial amino acid sequence similarity with an amino acid sequence of a CRSP. Natural variants include individual, polymorphic, allelic, strain, or species variants.

Amino acid sequence similarity, or sequence identity, is determined by optimizing residue matches, if necessary, by introducing gaps as required. This changes when considering conservative substitutions as matches. Physiocochemical conservative residue substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Homologous amino acid sequences include natural polymorphic, allelic, and interspecies variations in each respective protein sequence. Typical homologous proteins or peptides will have from 50-100% similarity (if gaps can be introduced), to 75-100% similarity (if conservative substitutions are included) with the amino acid sequence of the CRSP. Similarity measures will be at least about 50%, generally at least 60%, more generally at least 65%, usually at least 70%, more usually at least 75%, preferably at least 80%, and more preferably at least 80%, and in particularly preferred embodiments, at least 85% or more. See also Needleham, et al. (1970) J. Mol. Biol. 48:443-453; Sankoff, et al. (1983) Time Warps, String Edits, and Macromolecules: The

Theory and Practice of Sequence Comparison Chapter One, Addison-Wesley, Reading, MA; and software packages from IntelliGenetics, Mountain View, CA; and the University of Wisconsin Genetics Computer Group, Madison, WI.

5        For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and  
10       sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

15       Optical alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443, by the search for similarity  
20       method of Pearson and Lipman (1988) Proc. Nat'l Acad. Sci. USA 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual  
25       inspection (see generally Ausubel et al., supra).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It  
30       also plots a tree or dendrogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng and Doolittle (1987) J. Mol. Evol. 35:351-360. The method used is similar to the method described by Higgins  
35       and Sharp (1989) CABIOS 5:151-153. The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two

aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described Altschul, et al. (1990) J. Mol. Biol. 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul, et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the



sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) Proc. Nat'l Acad. Sci. USA 89:10915) alignments (B) of  
5 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see,  
10 e.g., Karlin and Altschul (1993) Proc. Nat'l Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability  
15 by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test  
nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most  
20 preferably less than about 0.001.

A further indication that two nucleic acid sequences of polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide  
25 encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are  
30 substantially identical is that the two molecules hybridize to each other under stringent conditions, as described below.

Nucleic acids encoding rodent CRSP proteins will often hybridize to the nucleic acid sequence of SEQ ID  
35 NO: 1, 3, 5, 7 or 9 under stringent conditions. Nucleic acids encoding primate CRSP proteins will often hybridize to the nucleic acid sequence of SEQ ID NO: 13 under stringent conditions. For example, nucleic acids encoding mouse CRSP proteins will normally hybridize to

the nucleic acid of SEQ ID NO: 1 under stringent hybridization conditions. Generally, stringent conditions are selected to be about 10° C lower than the thermal melting point (T<sub>m</sub>) for the probe sequence at a defined ionic strength and pH. The T<sub>m</sub> is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is about 0.2 molar at pH 7 and the temperature is at least about 50° C. Other factors may significantly affect the stringency of hybridization, including, among others, base composition and size of the complementary strands, the presence of organic solvents such as formamide, and the extent of base mismatching. A preferred embodiment will include nucleic acids which will bind to disclosed sequences in 50% formamide and 200 mM NaCl at 42° C.

An isolated CRSP nucleic acid sequence can be readily modified by nucleotide substitutions, nucleotide deletions, nucleotide insertions, and short inversions of nucleotide stretches. These modifications result in novel DNA sequences which encode CRSP antigens, their derivatives, or proteins having highly similar physiological, immunogenic, or antigenic activity.

Modified sequences can be used to produce mutant antigens or to enhance expression. Enhanced expression may involve gene amplification, increased transcription, increased translation, and other mechanisms. Such mutant CRSP derivatives include predetermined or site-specific mutations of the respective protein or its fragments. "Mutant CRSP" encompasses a polypeptide otherwise falling within the homology definition of the CRSP as set forth above, but having an amino acid sequence which differs from that of a CRSP as found in nature, whether by way of deletion, substitution, or insertion. In particular, "site specific mutant CRSP" generally includes proteins having significant similarity with a protein having a sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14, and as sharing various biological activities, e.g., antigenic or

immunogenic, with those sequences, and in preferred  
embodiments contain most or all of the disclosed  
sequence. This applies also to polymorphic variants from  
different individuals. Similar concepts apply to  
5 different CRSP proteins, particularly those found in  
various warm blooded animals, e.g., mammals and birds.  
As stated before, it is emphasized that descriptions are  
generally meant to encompass other CRSP proteins, not  
limited to the mouse embodiments specifically discussed.

10 Although site specific mutation sites are  
predetermined, mutants need not be site specific. CRSP  
mutagenesis can be conducted by making amino acid  
insertions or deletions. Substitutions, deletions,  
insertions, or any combinations may be generated to  
15 arrive at a final construct. Insertions may include  
amino- or carboxyl- terminal fusions, e.g., epitope tags.  
Random mutagenesis can be conducted at a target codon and  
the expressed mutants can then be screened for the  
desired activity. Methods for making substitution  
20 mutations at predetermined sites in DNA having a known  
sequence are well known in the art, e.g., by M13 primer  
mutagenesis or polymerase chain reaction (PCR)  
techniques. See also, Sambrook, et al. (1989) and  
Ausubel, et al. (1987 and Supplements). The mutations in  
25 the DNA normally should not place coding sequences out of  
reading frames and preferably will not create  
complementary regions that could hybridize to produce  
secondary mRNA structure such as loops or hairpins.

The present invention also provides recombinant  
30 proteins, e.g., heterologous fusion proteins using  
segments from these proteins. A heterologous fusion  
protein is a fusion of proteins or segments which are  
naturally not normally fused in the same manner. Thus,  
the fusion product of an immunoglobulin with a CRSP  
35 polypeptide is a continuous protein molecule having  
sequences fused in a typical peptide linkage, typically  
made as a single translation product and exhibiting  
properties derived from each source peptide. One  
preferred embodiment is fusion of an Ig domain to the

carboxy terminus of the protein. A similar concept applies to heterologous nucleic acid sequences.

In addition, new constructs may be made from combining similar functional domains from other proteins.

5 Protein-binding or other segments may be "swapped" between different new fusion polypeptides or fragments, e.g., different CRSP embodiments. See, e.g., Cunningham, et al. (1989) Science 243:1330-1336; and O'Dowd, et al. (1988) J. Biol. Chem. 263:15985-15992. Thus, new

10 chimeric polypeptides exhibiting new combinations of specificities will result from the functional linkage of protein-binding specificities and other functional domains.

15 VIII. Binding Agent:CRSP Protein Complexes

A CRSP protein that specifically binds to or that is specifically immunoreactive with an antibody generated against a defined immunogen, such as an immunogen consisting of the amino acid sequence of SEQ ID NO: 2, 4,

20 6, 8, 10, 12, or 14, is typically determined in an immunoassay. The immunoassay uses a polyclonal antiserum which was raised to a protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14, as appropriate. This antiserum is selected to have low crossreactivity against other

25 secreted proteins and any such crossreactivity is removed by immunoabsorbtion prior to use in the immunoassay.

In order to produce antisera for use in an immunoassay, the protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14, is isolated as described herein. For example,

30 recombinant protein may be produced in a mammalian cell line. An inbred strain of mice or rats, such as balb/c mice, is immunized with a protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14 using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol

35 (see Harlow and Lane, supra). Alternatively, a synthetic peptide, preferably near full length, derived from the sequences disclosed herein and conjugated to a carrier protein can be used an immunogen. Polyclonal sera are collected and titered against the immunogen protein in an

immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Polyclonal antisera with a titer of  $10^4$  or greater are selected and tested for their cross reactivity against, e.g., known proteins exhibiting sequence similarity such as cytokines or growth factors, using a competitive binding immunoassay such as the one described in Harlow and Lane, supra, at pages 570-573. Preferably two CRSPs are used in this determination in conjunction with various embodiments, e.g., mouse CRSPs.

Various forms of CRSPs are used to identify antibodies which are specifically bound. These proteins can be produced as recombinant proteins and isolated using standard molecular biology and protein chemistry techniques as described herein. Moreover, since the CRSPs seem to lack glycosylation sites, problems of post-translational modifications is lessened.

Immunoassays in the competitive binding format can be used for the crossreactivity determinations. For example, a protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14 can be immobilized to a solid support. Proteins added to the assay compete with the binding of the antisera to the immobilized antigen. The ability of the above proteins to compete with the binding of the antisera to the immobilized protein is compared to the protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14. The percent crossreactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% crossreactivity with each of the proteins listed above are selected and pooled. The cross-reacting antibodies are then removed from the pooled antisera by immunoabsorption with the above-listed proteins.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein to the immunogen protein (e.g., the CRSP cysteine rich motifs of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14). In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to

inhibit 50% of the binding of the antisera to the immobilized protein is determined. If, e.g., the amount of the second protein required is less than twice the amount of the protein of SEQ ID NO: 2 that is required, then the second protein is said to specifically bind to an antibody generated to the immunogen.

It is understood that CRSP proteins are a family of homologous proteins that comprise multiple genes. For a particular gene product, such as the C2 CRSP protein or the C23 CRSP protein, the term refers not only to the amino acid sequences disclosed herein, but also to other proteins that are polymorphic, allelic, non-allelic, or species variants. It is also understood that the term "mouse CRSP" includes nonnatural mutations introduced by deliberate mutation using conventional recombinant technology such as single site mutation, or by excising very short sections of DNA encoding CRSP proteins, or by substituting new amino acids, or adding new amino acids. Such minor alterations must substantially maintain a particular feature, e.g., the immunoidentity of the original molecule and/or a biological activity. Thus, these alterations include proteins that are specifically immunoreactive with a designated naturally occurring CRSP protein, for example, the C2 CRSP protein shown in SEQ ID NO: 2, or of SEQ ID NO: 4 or 6 or the C23 CRSP protein shown in SEQ ID NO: 14. The biological properties of the altered proteins can be determined by expressing the protein in an appropriate cell line and measuring, e.g., a chemotactic effect. Particular protein modifications considered minor would include conservative substitution of amino acids with similar chemical properties, as described above for the CRSP family as a whole. By aligning, e.g., a protein optimally with the protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, or 14, and by using the conventional immunoassays described herein to determine immunoidentity, or by using chemotaxis assays, one can determine the protein compositions of the invention.

## IX. Functional Variants

The blocking of physiological response to CRSPs may result from the inhibition of binding of the protein to its receptor, e.g., through competitive inhibition.

- 5 Thus, in vitro assays of the present invention will often use isolated protein, membranes from cells expressing a recombinant membrane associated CRSP, soluble fragments comprising receptor binding segments of these proteins, or fragments attached to solid phase substrates. These  
10 assays will also allow for the diagnostic determination of the effects of either binding segment mutations and modifications, or protein mutations and modifications, e.g., protein analogs. This invention also contemplates the use of competitive drug screening assays, e.g., where  
15 neutralizing antibodies to antigen or receptor fragments compete with a test compound for binding to the protein. In this manner, the antibodies can be used to detect the presence of a polypeptide which shares one or more antigenic binding sites of the protein and can also be  
20 used to occupy binding sites on the protein that might otherwise interact with a receptor.

"Derivatives" of CRSP antigens include amino acid sequence mutants, glycosylation variants, and covalent or aggregate conjugates with other chemical moieties.

- 25 Covalent derivatives can be prepared by linkage of functionalities to groups which are found in CRSP amino acid side chains or at the N- or C- termini, by means which are well known in the art. These derivatives can include, without limitation, aliphatic esters or amides  
30 of the carboxyl terminus, or of residues containing carboxyl side chains, O-acyl derivatives of hydroxyl group-containing residues, and N-acyl derivatives of the amino terminal amino acid or amino-group containing residues, e.g., lysine or arginine. Acyl groups are  
35 selected from the group of alkyl-moieties including C3 to C18 normal alkyl, thereby forming alkanoyl aroyl species. Covalent attachment to carrier proteins may be important when immunogenic moieties are haptens.

In particular, glycosylation alterations are included, e.g., made by modifying the glycosylation patterns of a polypeptide during its synthesis and processing, or in further processing steps. While the primary sequences suggest the absence of glycosylation sites, the sequences may be modified to incorporate such. Particularly preferred means for accomplishing this are by exposing the polypeptide to glycosylating enzymes derived from cells which normally provide such processing, e.g., mammalian glycosylation enzymes. Deglycosylation enzymes are also contemplated. Also embraced are versions of the same primary amino acid sequence which have other minor modifications, including phosphorylated amino acid residues, e.g., phosphotyrosine, phosphoserine, or phosphothreonine, or other moieties, including ribosyl groups or cross-linking reagents.

A major group of derivatives are covalent conjugates of the CRSP or fragments thereof with other proteins or polypeptides. These derivatives can be synthesized in recombinant culture such as N- or C-terminal fusions or by the use of agents known in the art for their usefulness in cross-linking proteins through reactive side groups. Preferred protein derivatization sites with cross-linking agents are at free amino groups, carbohydrate moieties, and cysteine residues.

Fusion polypeptides between CRSPs and other homologous or heterologous proteins are also provided. Many growth factors and cytokines are homodimeric entities, and a repeat construct may have various advantages, including lessened susceptibility to proteolytic degradation. Moreover, many receptors require dimerization to transduce a signal, and various dimeric proteins or domain repeats can be desirable. Heterologous polypeptides may be fusions between different surface markers, resulting in, e.g., a hybrid protein exhibiting receptor binding specificity. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities



of the derivative proteins. Typical examples are fusions of a reporter polypeptide, e.g., luciferase, with a segment or domain of a protein, e.g., a receptor-binding segment, so that the presence or location of the fused protein may be easily determined. See, e.g., Dull, et al., U.S. Patent No. 4,859,609. Other gene fusion partners include bacterial  $\beta$ -galactosidase, trpE, Protein A,  $\beta$ -lactamase, alpha amylase, alcohol dehydrogenase, and yeast alpha mating factor. See, e.g., Godowski, et al. (1988) Science 241:812-816.

Such polypeptides may also have amino acid residues which have been chemically modified by phosphorylation, sulfonation, biotinylation, or the addition or removal of other moieties, particularly those which have molecular shapes similar to phosphate groups. In some embodiments, the modifications will be useful labeling reagents, or serve as purification targets, e.g., affinity ligands.

This invention also contemplates the use of derivatives of CRSPs other than variations in amino acid sequence or glycosylation. Such derivatives may involve covalent or aggregative association with chemical moieties. These derivatives generally fall into the three classes: (1) salts, (2) side chain and terminal residue covalent modifications, and (3) adsorption complexes, for example with cell membranes. Such covalent or aggregative derivatives are useful as immunogens, as reagents in immunoassays, or in purification methods such as for affinity purification of ligands or other binding ligands. For example, a CRSP antigen can be immobilized by covalent bonding to a solid support such as cyanogen bromide-activated SEPHAROSE, by methods which are well known in the art, or adsorbed onto polyolefin surfaces, with or without glutaraldehyde cross-linking, for use in the assay or purification of anti-CRSP antibodies or its receptor. The CRSPs can also be labeled with a detectable group, e.g., radioiodinated by the chloramine T procedure, covalently bound to rare earth chelates, or conjugated to another fluorescent moiety for use in diagnostic assays. Purification of

CRSPs may be effected by immobilized antibodies or receptor.

Isolated CRSP genes will allow transformation of cells lacking expression of corresponding CRSPs, e.g., either species types or cells which lack corresponding proteins and exhibit negative background activity. Expression of transformed genes will allow isolation of antigenically pure cell lines, with defined or single specie variants. This approach will allow for more sensitive detection and discrimination of the physiological effects of CRSP receptor proteins. Subcellular fragments, e.g., cytoplasts or membrane fragments, can be isolated and used.

#### 15 X. Uses

The present invention provides reagents which will find use in diagnostic applications as described elsewhere herein, e.g., in the general description for developmental abnormalities, or below in the description of kits for diagnosis. Each of these embodiments of the family are associated rather specifically with an inflammatory or immunologically active tissue.

CRSP nucleotides, e.g., DNA or RNA, may be used as a component in a diagnostic assay. For instance, the nucleotide sequences provided may be labeled using, e.g., <sup>32</sup>p or biotin and used to probe standard restriction fragment polymorphism blots, providing a measurable character to aid in distinguishing between individuals. Such probes may be used in well-known forensic techniques such as genetic fingerprinting. In addition, nucleotide probes made from CRSP sequences may be used in in situ assays to detect chromosomal abnormalities. For instance, rearrangements in the mouse chromosome encoding a CRSP gene may be detected via well-known in situ techniques, using CRSP probes in conjunction with other known chromosome markers.

Antibodies and other binding agents directed towards CRSP proteins or nucleic acids may be used to purify the corresponding CRSP molecule. As described in the

Examples below, antibody purification of CRSP components is both possible and practicable. Antibodies and other binding agents may also be used in a diagnostic fashion to determine whether CRSP components are present in a tissue sample or cell population using well-known techniques described herein. Specific medical conditions correlating with expression of the respective embodiments is described. The ability to attach a binding agent to a CRSP provides a means to diagnose disorders associated with CRSP misregulation. Antibodies and other CRSP binding agents may also be useful as histological markers. As described in the examples below, CRSP expression is limited to specific tissue types. By directing a probe, such as an antibody or nucleic acid to a CRSP it is possible to use the probe to distinguish tissue and cell types in situ or in vitro.

This invention also provides reagents with significant therapeutic value. The CRSPs (naturally occurring or recombinant), fragments thereof, and antibodies thereto, along with compounds identified as having binding affinity to a CRSP, are useful in the treatment of conditions associated with abnormal physiology or development, including abnormal proliferation, e.g., inflammatory conditions, cancerous conditions, or degenerative conditions. Abnormal proliferation, regeneration, degeneration, and atrophy may be modulated by appropriate therapeutic treatment using the compositions provided herein. For example, a disease or disorder associated with abnormal expression or abnormal signaling by a CRSP is a target for an agonist or antagonist of the protein. The proteins likely play a role in regulation or development of various cells, e.g., lymphoid cells, which affect immunological responses.

Other abnormal developmental conditions are known in cell types shown to possess CRSP mRNA by northern blot analysis. See Berkow (ed.) The Merck Manual of Diagnosis and Therapy, Merck & Co., Rahway, NJ; and Thorn, et al. Harrison's Principles of Internal Medicine, McGraw-Hill,

NY. Developmental or functional abnormalities, e.g., of the immune system, cause significant medical abnormalities and conditions which may be susceptible to prevention or treatment using compositions provided  
5 herein. The role of epithelial cells in such conditions may be important.

Recombinant CRSP or CRSP antibodies can be purified and then administered to a patient. These reagents can be combined for therapeutic use with additional active or  
10 inert ingredients, e.g., in conventional pharmaceutically acceptable carriers or diluents, e.g., immunogenic adjuvants, along with physiologically innocuous stabilizers and excipients. These combinations can be  
15 lyophilization in dosage vials or storage in stabilized aqueous preparations. This invention also contemplates use of antibodies or binding fragments thereof, including forms which are not complement binding. In particular, the C2 proteins seem to be associated with lung  
20 physiology, and combination thereof with other lung related therapeutics is suggested; the C18 proteins associate with colon physiology and combination thereof with other colon related therapeutics is suggested; the C19 proteins seem to be associated with joint or  
25 arthritic physiology, and combination thereof with other joint related therapeutics is suggested; and the C10 seem to be associated with lung or colon physiology, and combination thereof with other lung or colon related therapeutics is suggested. See, e.g., Berkow (ed.) The  
30 Merck Manual of Diagnosis and Therapy, Merck & Co., Rahway, NJ; and Thorn, et al. Harrison's Principles of Internal Medicine, McGraw-Hill, NY.

Drug screening using antibodies or receptor or fragments thereof can identify compounds having binding  
35 affinity to CRSPs, including isolation of associated components. Subsequent biological assays can then be utilized to determine if the compound has intrinsic stimulating activity and is therefore a blocker or antagonist in that it blocks the activity of the protein.

Likewise, a compound having intrinsic stimulating activity can activate the receptor and is thus an agonist in that it simulates the activity of a CRSP. This invention further contemplates the therapeutic use of antibodies to CRSPs as antagonists. This approach should be particularly useful with other CRSP species variants.

The quantities of reagents necessary for effective therapy will depend upon many different factors, including means of administration, target site, physiological state of the patient, and other medicants administered. Thus, treatment dosages should be titrated to optimize safety and efficacy. Typically, dosages used in vitro may provide useful guidance in the amounts useful for in situ administration of these reagents.

Animal testing of effective doses for treatment of particular disorders will provide further predictive indication of human dosage. Various considerations are described, e.g., in Gilman, et al. (eds. 1990) Goodman and Gilman's: The Pharmacological Bases of Therapeutics (8th ed.) Pergamon Press; and (1990) Remington's Pharmaceutical Sciences (17th ed.) Mack Publishing Co., Easton, PA. Methods for administration are discussed therein and below, e.g., for oral, intravenous, intraperitoneal, or intramuscular administration, transdermal diffusion, and others. Pharmaceutically acceptable carriers will include water, saline, buffers, and other compounds described, e.g., in the Merck Index, Merck & Co., Rahway, NJ. Dosage ranges would ordinarily be expected to be in amounts lower than 1 mM concentrations, typically less than about 10  $\mu$ M concentrations, usually less than about 100 nM, preferably less than about 10 pM (picomolar), and most preferably less than about 1 fM (femtomolar), with an appropriate carrier. Slow release formulations, or a slow release apparatus will often be utilized for continuous administration.

CRSPs, fragments thereof, and antibodies to it or its fragments, antagonists, and agonists, may be administered directly to the host to be treated or,

depending on the size of the compounds, it may be desirable to conjugate them to carrier proteins such as ovalbumin or serum albumin prior to their administration. Therapeutic formulations may be administered in any conventional dosage formulation. While it is possible for the active ingredient to be administered alone, it is preferable to present it as a pharmaceutical formulation. Formulations typically comprise at least one active ingredient, as defined above, together with one or more acceptable carriers thereof. Each carrier should be both pharmaceutically and physiologically acceptable in the sense of being compatible with the other ingredients and not injurious to the patient. Formulations include those suitable for oral, rectal, nasal, or parenteral (including subcutaneous, intramuscular, intravenous and intradermal) administration. The formulations may conveniently be presented in unit dosage form and may be prepared by any methods well known in the art of pharmacy. See, e.g., Gilman, et al. (eds. 1990) Goodman and Gilman's: The Pharmacological Bases of Therapeutics (8th ed.) Pergamon Press; and (1990) Remington's Pharmaceutical Sciences (17th ed.) Mack Publishing Co., Easton, PA; Avis, et al. (eds. 1993) Pharmaceutical Dosage Forms: Parenteral Medications Dekker, NY; Lieberman, et al. (eds. 1990) Pharmaceutical Dosage Forms: Tablets Dekker, NY; and Lieberman, et al. (eds. 1990) Pharmaceutical Dosage Forms: Disperse Systems Dekker, NY. The therapy of this invention may be combined with or used in association with other therapeutic agents.

Both the naturally occurring and the recombinant forms of the CRSPs of this invention are particularly useful in kits and assay methods which are capable of screening compounds for binding activity to the proteins. Several methods of automating assays have been developed in recent years so as to permit screening of tens of thousands of compounds in a short period. See, e.g., Fodor, et al. (1991) Science 251:767-773, and other descriptions of chemical diversity libraries, which

describe means for testing of binding affinity by a plurality of compounds. The development of suitable assays can be greatly facilitated by the availability of large amounts of purified, soluble CRSP as provided by this invention.

For example, antagonists can normally be found once the protein has been structurally defined. Testing of potential protein analogs is now possible upon the development of highly automated assay methods using a purified receptor. In particular, new agonists and antagonists will be discovered by using screening techniques described herein. Of particular importance are compounds found to have a combined binding affinity for multiple CRSP receptors, e.g., compounds which can serve as antagonists for species variants of a CRSP.

This invention is particularly useful for screening compounds by using recombinant protein in a variety of drug screening techniques. The advantages of using a recombinant protein in screening for specific ligands include: (a) improved renewable source of the CRSP from a specific source; (b) potentially greater number of ligands per cell giving better signal to noise ratio in assays; and (c) species variant specificity (theoretically giving greater biological and disease specificity).

One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant DNA molecules expressing a CRSP receptor. Cells may be isolated which express a receptor in isolation from any others. Such cells, either in viable or fixed form, can be used for standard ligand/receptor binding assays. See also, Parce, et al. (1989) Science 246:243-247; and Owicki, et al. (1990) Proc. Nat'l Acad. Sci. USA 87:4007-4011, which describe sensitive methods to detect cellular responses. Competitive assays are particularly useful, where the cells (source of CRSP) are contacted and incubated with a labeled receptor or antibody having known binding affinity to the ligand, such as  $^{125}\text{I}$ -antibody, and a test sample whose binding

affinity to the binding composition is being measured. The bound and free labeled binding compositions are then separated to assess the degree of ligand binding. The amount of test compound bound is inversely proportional to the amount of labeled receptor binding to the known source. Any one of numerous techniques can be used to separate bound from free ligand to assess the degree of ligand binding. This separation step could typically involve a procedure such as adhesion to filters followed by washing, adhesion to plastic followed by washing, or centrifugation of the cell membranes. Viable cells could also be used to screen for the effects of drugs on CRSP mediated functions, e.g., second messenger levels, i.e.,  $Ca^{++}$ ; cell proliferation; inositol phosphate pool changes; and others. Some detection methods allow for elimination of a separation step, e.g., a proximity sensitive detection system. Calcium sensitive dyes will be useful for detecting  $Ca^{++}$  levels, with a fluorimeter or a fluorescence cell sorting apparatus.

Another method utilizes membranes from transformed eukaryotic or prokaryotic host cells as the source of a CRSP. These cells are stably transformed with DNA vectors directing the expression of a CRSP, e.g., an engineered membrane bound form. Essentially, the membranes would be prepared from the cells and used in a receptor/ligand binding assay such as the competitive assay set forth above.

Still another approach is to use solubilized, unpurified or solubilized, purified CRSP from transformed eukaryotic or prokaryotic host cells. This allows for a "molecular" binding assay with the advantages of increased specificity, the ability to automate, and high drug test throughput.

Another technique for drug screening involves an approach which provides high throughput screening for compounds having suitable binding affinity to a CRSP antibody and is described in detail in Geysen, European Patent Application 84/03564, published on September 13, 1984. First, large numbers of different small peptide



test compounds are synthesized on a solid substrate, e.g., plastic pins or some other appropriate surface, see Fodor, et al., supra. Then all the pins are reacted with solubilized, unpurified or solubilized, purified CRSP antibody, and washed. The next step involves detecting bound CRSP antibody.

Rational drug design may also be based upon structural studies of the molecular shapes of the CRSP and other effectors or analogs. See, e.g., Methods in Enzymology vols. 202 and 203. Effectors may be other proteins which mediate other functions in response to ligand binding, or other proteins which normally interact with the receptor. One means for determining which sites interact with specific other proteins is a physical structure determination, e.g., x-ray crystallography or 2 dimensional NMR techniques. These will provide guidance as to which amino acid residues form molecular contact regions. For a detailed description of protein structural determination, see, e.g., Blundell and Johnson (1976) Protein Crystallography Academic Press, NY.

A purified CRSP can be coated directly onto plates for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to these ligands can be used as capture antibodies to immobilize the respective ligand on the solid phase.

#### XI. Kits

This invention also contemplates use of CRSPs, fragments thereof, peptides, and their fusion products in a variety of diagnostic kits and methods for detecting the presence of CRSP or a CRSP receptor. Typically the kit will have a compartment containing either a defined CRSP peptide or gene segment or a reagent which recognizes one or the other, e.g., receptor fragments or antibodies.

A kit for determining the binding affinity of a test compound to a CRSP would typically comprise a test compound; a labeled compound, e.g., a receptor or antibody having known binding affinity for the CRSP; a

source of CRSP (naturally occurring or recombinant); and a means for separating bound from free labeled compound, such as a solid phase for immobilizing the CRSP. Once compounds are screened, those having suitable binding affinity to the CRSP can be evaluated in suitable biological assays, as are well known in the art, to determine whether they act as agonists or antagonists to the receptor. The availability of recombinant CRSP polypeptides also provide well defined standards for calibrating such assays.

A preferred kit for determining the concentration of, for example, a CRSP in a sample would typically comprise a labeled compound, e.g., receptor or antibody, having known binding affinity for the CRSP, a source of CRSP (naturally occurring or recombinant), and a means for separating the bound from free labeled compound, for example, a solid phase for immobilizing the CRSP. Compartments containing reagents, and instructions, will normally be provided.

Antibodies, including antigen binding fragments, specific for the CRSP or ligand fragments are useful in diagnostic applications to detect the presence of elevated levels of CRSP and/or its fragments. Such diagnostic assays can employ lysates, live cells, fixed cells, immunofluorescence, cell cultures, body fluids, and further can involve the detection of antigens related to the ligand in a body fluid, e.g., serum, or the like. Diagnostic assays may be homogeneous (without a separation step between free reagent and antigen-CRSP complex) or heterogeneous (with a separation step). Various commercial assays exist, such as radioimmunoassay (RIA), enzyme-linked immunosorbent assay (ELISA), enzyme immunoassay (EIA), enzyme-multiplied immunoassay technique (EMIT), substrate-labeled fluorescent immunoassay (SLFIA), and the like. For example, unlabeled antibodies can be employed by using a second antibody which is labeled and which recognizes the antibody to a CRSP or to a particular fragment thereof. Similar assays have also been extensively discussed in

the literature. See, e.g., Harlow and Lane (1988) Antibodies: A Laboratory Manual, CSH Press, NY; Chan (ed. 1987) Immunoassay: A Practical Guide Academic Press, Orlando, FL; Price and Newman (eds. 1991) Principles and Practice of Immunoassay Stockton Press, NY; and Ngo (ed. 1988) Nonisotopic Immunoassay Plenum Press, NY.

Anti-idiotypic antibodies may have similar use to diagnose presence of antibodies against a CRSP, as such may be diagnostic of various abnormal states. For example, overproduction of CRSP may result in production of various immunological or other medical reactions which may be diagnostic of abnormal physiological states, e.g., in cell growth, activation, or differentiation.

Frequently, the reagents for diagnostic assays are supplied in kits, so as to optimize the sensitivity of the assay. For the subject invention, depending upon the nature of the assay, the protocol, and the label, either labeled or unlabeled antibody or receptor, or labeled CRSP is provided. This is usually in conjunction with other additives, such as buffers, stabilizers, materials necessary for signal production such as substrates for enzymes, and the like. Preferably, the kit will also contain instructions for proper use and disposal of the contents after use. Typically the kit has compartments for each useful reagent. Desirably, the reagents are provided as a dry lyophilized powder, where the reagents may be reconstituted in an aqueous medium providing appropriate concentrations of reagents for performing the assay.

Many of the aforementioned constituents of the drug screening and the diagnostic assays may be used without modification, or may be modified in a variety of ways. For example, labeling may be achieved by covalently or non-covalently joining a moiety which directly or indirectly provides a detectable signal. In any of these assays, the protein, test compound, CRSP, or antibodies thereto can be labeled either directly or indirectly. Possibilities for direct labeling include label groups: radiolabels such as  $^{125}\text{I}$ , enzymes (U.S. Pat. No.

3,645,090) such as peroxidase and alkaline phosphatase, and fluorescent labels (U.S. Pat. No. 3,940,475) capable of monitoring the change in fluorescence intensity, wavelength shift, or fluorescence polarization.

- 5 Possibilities for indirect labeling include biotinylation of one constituent followed by binding to avidin coupled to one of the above label groups.

There are also numerous methods of separating the bound from the free ligand, or alternatively the bound  
10 from the free test compound. The CRSP can be immobilized on various matrices followed by washing. Suitable matrices include plastic such as an ELISA plate, filters, and beads. Methods of immobilizing the CRSP to a matrix include, without limitation, direct adhesion to plastic,  
15 use of a capture antibody, chemical coupling, and biotin-avidin. The last step in this approach involves the precipitation of ligand/receptor or ligand/antibody complex by any of several methods including those utilizing, e.g., an organic solvent such as polyethylene glycol or a salt such as ammonium sulfate. Other  
20 suitable separation techniques include, without limitation, the fluorescein antibody magnetizable particle method described in Rattle, et al. (1984) Clin. Chem. 30:1457-1461, and the double antibody magnetic  
25 particle separation as described in U.S. Pat. No. 4,659,678.

- Methods for linking proteins or their fragments to the various labels have been extensively reported in the literature and do not require detailed discussion here.
- 30 Many of the techniques involve the use of activated carboxyl groups either through the use of carbodiimide or active esters to form peptide bonds, the formation of thioethers by reaction of a mercapto group with an activated halogen such as chloroacetyl, or an activated  
35 olefin such as maleimide, for linkage, or the like. Fusion proteins will also find use in these applications.

Another diagnostic aspect of this invention involves use of oligonucleotide or polynucleotide sequences taken from the sequence of a CRSP. These sequences can be used

as probes for detecting levels of the CRSP message in samples from natural sources, or patients suspected of having an abnormal condition, e.g., cancer or developmental problem. The preparation of both RNA and DNA nucleotide sequences, the labeling of the sequences, and the preferred size of the sequences has received ample description and discussion in the literature. Normally an oligonucleotide probe should have at least about 14 nucleotides, usually at least about 18 nucleotides, and the polynucleotide probes may be up to several kilobases. Various labels may be employed, most commonly radionuclides, particularly  $^{32}\text{P}$ . However, other techniques may also be employed, such as using biotin modified nucleotides for introduction into a polynucleotide. The biotin then serves as the site for binding to avidin or antibodies, which may be labeled with a wide variety of labels, such as radionuclides, fluorophores, enzymes, or the like. Alternatively, antibodies may be employed which can recognize specific duplexes, including DNA duplexes, RNA duplexes, DNA-RNA hybrid duplexes, or DNA-protein duplexes. The antibodies in turn may be labeled and the assay carried out where the duplex is bound to a surface, so that upon the formation of duplex on the surface, the presence of antibody bound to the duplex can be detected. The use of probes to the novel anti-sense RNA may be carried out using many conventional techniques such as nucleic acid hybridization, plus and minus screening, recombinational probing, hybrid released translation (HRT), and hybrid arrested translation (HART). This also includes amplification techniques such as polymerase chain reaction (PCR).

Diagnostic kits which also test for the qualitative or quantitative presence of other markers are also contemplated. Diagnosis or prognosis may depend on the combination of multiple indications used as markers. Thus, kits may test for combinations of markers. See, e.g., Viallet, et al. (1989) Progress in Growth Factor Res. 1:89-97.

## XII. Receptor Isolation

Having isolated a binding partner of a specific interaction, methods exist for isolating the counter-partner. See, Gearing, et al. (1989) EMBO J. 8:3667-3676. For example, means to label a CRSP without interfering with the binding to its receptor can be determined. For example, an affinity label or epitope tag can be fused to either the amino- or carboxyl-terminus of the ligand. An expression library can be screened for specific binding of the CRSP, e.g., by cell sorting, or other screening to detect subpopulations which express such a binding component. See, e.g., Ho, et al. (1993) Proc. Nat'l Acad. Sci. USA 90:11267-11271. Alternatively, a panning method may be used. See, e.g., Seed and Aruffo (1987) Proc. Nat'l Acad. Sci. USA 84:3365-3369. A two-hybrid selection system may also be applied making appropriate constructs with the available CRSP sequences. See, e.g., Fields and Song (1989) Nature 340:245-246.

Protein cross-linking techniques with label can be applied to isolate binding partners of a CRSP. This would allow identification of proteins which specifically interact with a CRSP, e.g., in a ligand-receptor like manner. It is likely that the receptor will be found by expression in a system which is capable of expressing a membrane protein in a form capable of exhibiting ligand binding capability.

The broad scope of this invention is best understood with reference to the following examples, which are not intended to limit the invention to specific embodiments.

## EXAMPLES

### I. General Methods

Many of the standard methods below are described or referenced, e.g., in Maniatis, et al. (1982) Molecular Cloning, A Laboratory Manual Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY; Sambrook, et

- al. (1989) Molecular Cloning: A Laboratory Manual (2d ed.) Vols. 1-3, CSH Press, NY; Ausubel, et al., Biology Greene Publishing Associates, Brooklyn, NY; or Ausubel, et al. (1987 and Supplements) Current Protocols in  
5 Molecular Biology Wiley/Greene, NY; Innis, et al. (eds. 1990) PCR Protocols: A Guide to Methods and Applications Academic Press, NY. Methods for protein purification include such methods as ammonium sulfate precipitation, column chromatography, electrophoresis, centrifugation,  
10 crystallization, and others. See, e.g., Ausubel, et al. (1987 and periodic supplements); Deutscher (1990) "Guide to Protein Purification," Methods in Enzymology vol. 182, and other volumes in this series; and manufacturer's literature on use of protein purification products, e.g.,  
15 Pharmacia, Piscataway, NJ, or Bio-Rad, Richmond, CA. Combination with recombinant techniques allow fusion to appropriate segments (epitope tags), e.g., to a FLAG sequence or an equivalent which can be fused, e.g., via a protease-removable sequence. See, e.g., Hochuli (1989)  
20 Chemische Industrie 12:69-70; Hochuli (1990) "Purification of Recombinant Proteins with Metal Chelate Absorbent" in Setlow (ed.) Genetic Engineering, Principle and Methods 12:87-98, Plenum Press, NY; and Crowe, et al. (1992) OIAexpress: The High Level Expression & Protein  
25 Purification System QUIAGEN, Inc., Chatsworth, CA.

Standard immunological techniques are described, e.g., in Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; and Methods in Enzymology volumes. 70, 73, 74, 84, 92, 93, 108, 116, 121, 132, 150, 162, and  
30 163. Assays for neural cell biological activities are described, e.g., in Wouterlood (ed. 1995) Neuroscience Protocols modules 10, Elsevier; Methods in Neurosciences Academic Press; and Neuromethods Humana Press, Totowa, NJ. Methodology of developmental systems is described,  
35 e.g., in Meisami (ed.) Handbook of Human Growth and Developmental Biology CRC Press; and Chrispeels (ed.) Molecular Techniques and Approaches in Developmental Biology Interscience. Defensin assays are well known,

and described, e.g., in Harwig, et al. (1994) Methods in Enzymology 236:160-172.

FACS analyses are described in Melamed, et al. (1990) Flow Cytometry and Sorting Wiley-Liss, Inc., New York, NY; Shapiro (1988) Practical Flow Cytometry Liss, New York, NY; and Robinson, et al. (1993) Handbook of Flow Cytometry Methods Wiley-Liss, New York, NY.

## II. cDNA libraries

10 RNA is carefully extracted from the appropriate source to maintain full length integrity of the messenger. cDNA is made from appropriate cell types and fetal tissues, e.g., from 2 µg of high quality mRNA. Typically, each library was typically quality controlled  
15 by three criteria: 1) Alkaline gel analysis of the first strand synthesis to evaluate size range of cDNA from > 0.5 - 5 kb, indicating high quality RNA and a good predictor of large insert sizes in the final library. 2) Upon ligation of cDNA, the number of independent clones  
20 was greater than  $1 \times 10^6$  clones before amplification. 3) Sequence analysis of randomly selected clones from each library typically revealed a high proportion of full length clones, and only very low levels of genomic or ribosomal RNA contamination (< 5%). It was found that,  
25 although standard RNA blot analysis of gene expression levels is somewhat more sensitive, a positive signal in a cDNA library is roughly comparable to RNA analysis, particularly in judging the presence or absence of a particular gene in a certain cell type or tissue. Large  
30 scale plasmid DNA preparation of amplified libraries was performed, e.g., using a Giga prep (Qiagen, Chatsworth, CA).

Southern Analysis: DNA (5 µg) from the primary amplified cDNA library was digested with appropriate  
35 restriction enzymes to release the inserts, run on a 1% agarose gel and transferred to a nylon membrane (Schleicher and Schuell, Keene, NH).

Samples for mRNA isolation include: resting mouse fibroblastic L cell line (C200); Braf:ER (Braf fusion to



estrogen receptor) transfected cells, control (C201); T cells, TH1 polarized (Mel14 bright, CD4+ cells from spleen, polarized for 7 days with IFN- $\gamma$  and anti IL-4; T200); T cells, TH2 polarized (Mel14 bright, CD4+ cells from spleen, polarized for 7 days with IL-4 and anti-IFN- $\gamma$ ; T201); T cells, highly TH1 polarized (see Openshaw, et al. (1995) J. Exp. Med. 182:1357-1367; activated with anti-CD3 for 2, 6, 16 h pooled; T202); T cells, highly TH2 polarized (see Openshaw, et al. (1995) J. Exp. Med. 182:1357-1367; activated with anti-CD3 for 2, 6, 16 h pooled; T203); CD44- CD25+ pre T cells, sorted from thymus (T204); TH1 T cell clone D1.1, resting for 3 weeks after last stimulation with antigen (T205); TH1 T cell clone D1.1, 10  $\mu$ g/ml ConA stimulated 15 h (T206); TH2 T cell clone CDC35, resting for 3 weeks after last stimulation with antigen (T207); TH2 T cell clone CDC35, 10  $\mu$ g/ml ConA stimulated 15 h (T208); Mel14+ naive T cells from spleen, resting (T209); Mel14+ T cells, polarized to Th1 with IFN- $\gamma$ /IL-12/anti-IL-4 for 6, 12, 24 h pooled (T210); Mel14+ T cells, polarized to Th2 with IL-4/anti-IFN- $\gamma$  for 6, 13, 24 h pooled (T211); unstimulated mature B cell leukemia cell line A20 (B200); unstimulated B cell line CH12 (B201); unstimulated large B cells from spleen (B202); B cells from total spleen, LPS activated (B203); metrizamide enriched dendritic cells from spleen, resting (D200); dendritic cells from bone marrow, resting (D201); monocyte cell line RAW 264.7 activated with LPS 4 h (M200); bone-marrow macrophages derived with GM and M-CSF (M201); macrophage cell line J774, resting (M202); macrophage cell line J774 + LPS + anti-IL-10 at 0.5, 1, 3, 6, 12 h pooled (M203); macrophage cell line J774 + LPS + IL-10 at 0.5, 1, 3, 5, 12 h pooled (M204); aerosol challenged mouse lung tissue, Th2 primed, aerosol OVA challenge 7, 14, 23 h pooled (see Garlisi, et al. (1995) Clinical Immunology and Immunopathology 75:75-83; X206); Nippostrongylus-infected lung tissue (see Coffman, et al. (1989) Science 245:308-310; X200); total adult lung, normal (O200); total lung, rag-1 (see Schwarz, et al. (1993) Immunodeficiency 4:249-

252; O205); IL-10 K.O. spleen (see Kuhn, et al. (1991)  
Cell 75:263-274; X201); total adult spleen, normal  
(O201); total spleen, rag-1 (O207); IL-10 K.O. Peyer's  
patches (O202); total Peyer's patches, normal (O210); IL-  
5 10 K.O. mesenteric lymph nodes (X203); total mesenteric  
lymph nodes, normal (O211); IL-10 K.O. colon (X203);  
total colon, normal (O212); NOD mouse pancreas (see  
Makino, et al. (1980) Jikken Dobutsu 29:1-13; X205);  
total thymus, rag-1 (O208); total kidney, rag-1 (O209);  
10 total heart, rag-1 (O202); total brain, rag-1 (O203);  
total testes, rag-1 (O204); total liver, rag-1 (O206);  
rat normal joint tissue (O300); and rat arthritic joint  
tissue (X300).

### 15 III. Isolation of C2 CRSP embodiment

The original EST is from a mouse Nippo lung and was  
identified as being interesting and recognizing a  
potential leader sequence. This EST is represented twice  
in the subtraction library and once in the random  
20 sequence from the non-subtracted library (no matches in  
GenBank nt or nr, or dbest). Analysis of the OVA  
challenged lung revealed an additional two EST's in the  
random sequencing and 1 from the subtraction. ORF  
analysis identified a protein of about 111 residues.  
25 This ORF appears to have a cleavable signal peptide, with  
the predicted cleavage site at position 23-24. Using  
this protein sequence two additional mouse family members  
have been identified by TBLASTn searches. The features  
of this novel protein family include- ORF between 106-115  
30 amino acids, predicted N-terminal signal peptide,  
conserved pattern of cysteine residues (10 for the C2  
embodiments, 11 for the C18, C19, C10, and C23) in the  
mature protein, lack of N-linked glycosylation,  
restricted expression pattern correlated with disease  
35 states. The structure of this family shares similarity  
with members of the EGF family.

A clone encoding the C2 CRSP is isolated from a  
natural source by many different possible methods. Given  
the sequences provided herein, PCR primers or

hybridization probes are selected and/or constructed to isolate either genomic DNA segments or cDNA reverse transcripts. Appropriate cell sources include lung tissues. Genetic and polymorphic or allelic variants are  
5 isolated by screening a population of individuals, e.g., other strains of mice, other rodents, etc.

PCR based detection is performed by standard methods, preferably using primers from opposite ends of the coding sequence, but flanking segments might be  
10 selected for specific purposes.

Alternatively, hybridization probes are selected. Particular AT or GC contents of probes are selected depending upon the expected homology and mismatching expected. Appropriate stringency conditions are selected  
15 to balance an appropriate positive signal to background ratio. Successive washing steps are used to collect clones of greater homology.

Further clones are isolated using an antibody based selection procedure. Standard expression cloning methods  
20 are applied including, e.g., FACS staining of membrane associated expression product. The antibodies are used to identify clones producing a recognized protein. Alternatively, antibodies are used to purify a specific CRSP, with protein sequencing and standard means to  
25 isolate a gene encoding that protein.

Genomic sequence based methods will also allow for identification of sequences naturally available, or otherwise, which exhibit homology to the provided sequences.  
30

#### IV. Isolation of other rodent CRSP clones

Similar methods are used as above to isolate an appropriate mouse CRSP gene. Similar source materials as indicated above are used to isolate natural genes,  
35 including genetic, polymorphic, allelic, or strain variants. Species variants are also isolated using similar methods.

Another embodiment of a closely related C2 gene from mouse, designated C2b, is described in Table 1. Regions

of identity and divergence from the original C2 gene are apparent from Table 5.

A rat gene closely related to mouse C19 was also isolated. See Table 3.

5       Antigenic methods may be used to immunoprecipitate related molecules. In particular, native proteins likely share tertiary structure, as the cysteines are both numerous and conserved, and will likely cross react.

10       Additional data suggests that the genes encoding the various embodiments may be genetically linked. Isolation of the genomic region encoding one embodiment is likely to lead, using appropriate methods of chromosome walking or other methods, to the genes of other embodiments. PCR methodology may make use of the highly conserved segments  
15 of sequence, as indicated above. Particularly useful PCR primers spanning those conserved regions include, for the coding strand, TGT GGC THY GSC TGT GGM TCK TGG, and for the non-coding strand, CA GCA GCG SGC WSH KGT CCA GTC (SEQ ID NO: 15 and 16). These highly conserved primer  
20 segments are likely to be also useful to isolate species counterparts.

#### V. Isolation of C23 CRSP embodiment

25       A clone encoding the human C23 CRSP is isolated from a natural source by many different possible methods. Given the sequences provided herein, PCR primers or hybridization probes are selected and/or constructed to isolate either genomic DNA segments or cDNA reverse transcripts. Appropriate cell sources include inflamed  
30 tonsil tissue. Genetic and polymorphic or allelic variants are isolated by screening a population of individuals, e.g., other primates, etc.

PCR based detection is performed by standard methods, preferably using primers from opposite ends of  
35 the coding sequence, but flanking segments might be selected for specific purposes.

Alternatively, hybridization probes are selected. Particular AT or GC contents of probes are selected depending upon the expected homology and mismatching

expected. Appropriate stringency conditions are selected to balance an appropriate positive signal to background ratio. Successive washing steps are used to collect clones of greater homology.

- 5 Further clones are isolated using an antibody based selection procedure. Standard expression cloning methods are applied including, e.g., FACS staining of membrane associated expression product. The antibodies are used to identify clones producing a recognized protein.
- 10 Alternatively, antibodies are used to purify a specific CRSP, with protein sequencing and standard means to isolate a gene encoding that protein.

Genomic sequence based methods will also allow for identification of sequences naturally available, or otherwise, which exhibit homology to the provided sequences.

15

#### VI. Isolation of other primate CRSP clones

- Similar methods are used as above to isolate an appropriate primate CRSP gene. Similar source materials as indicated above are used to isolate natural genes, including genetic, polymorphic, allelic, or strain variants. Species variants are also isolated using similar methods.
- 20

- 25 Antigenic methods may be used to immunoprecipitate related molecules. In particular, native proteins likely share tertiary structure, as the cysteines are both numerous and conserved, and will likely cross react.

- Additional data suggests that the rodent genes encoding the various embodiments may be genetically linked. Isolation of the genomic region encoding one embodiment is likely to lead, using appropriate methods of chromosome walking or other methods, to the genes of other embodiments. PCR methodology may make use of the highly conserved segments of sequence, as indicated above. Particularly useful PCR primers spanning those conserved regions include, for the coding strand, TGT GGC THY GSC TGT GGM TCK TGG, and for the non-coding strand, CA GCA GCG SGC WSH KGT CCA GTC (SEQ ID NO: 15 and 16).
- 30
- 35

These highly conserved primer segments are likely to be also useful to isolate species counterparts.

#### VII. Expression; purification; characterization

5       With an appropriate clone from above, the coding sequence is inserted into an appropriate expression vector. This may be in a vector specifically selected for a prokaryote, yeast, insect, or higher vertebrate, e.g., mammalian expression system. Standard methods are  
10       applied to produce the gene product, preferably as a soluble secreted molecule, but will, in certain instances, also be made as an intracellular protein. Intracellular proteins typically require cell lysis to recover the protein, and insoluble inclusion bodies are a  
15       common starting material for further purification. Particularly preferred methods are to express in a eukaryotic, e.g., mouse cell. The protein has been well expressed and secreted.

      With a clone encoding a mouse CRSP, recombinant  
20       production means are used, although natural forms may be purified from appropriate sources. The protein product is purified by standard methods of protein purification, in certain cases, e.g., coupled with immunoaffinity methods. Immunoaffinity methods are used either as a  
25       purification step, as described above, or as a detection assay to determine the separation properties of the protein.

      Preferably, the protein is secreted into the medium, and the soluble product is purified from the medium in a  
30       soluble form. Alternatively, as described above, inclusion bodies from prokaryotic expression systems are a useful source of material. Typically, the insoluble protein is solubilized from the inclusion bodies and refolded using standard methods. Purification methods  
35       are developed as described above.

      The product of the purification method described above is characterized to determine many structural features. Standard physical methods are applied, e.g., amino acid analysis and protein sequencing. The

resulting protein is subjected to CD spectroscopy and other spectroscopic methods, e.g., NMR, ESR, mass spectroscopy, etc. The product is characterized to determine its molecular form and size, e.g., using gel chromatography and similar techniques. Understanding of the chromatographic properties will lead to more gentle or efficient purification methods. Preliminary analysis suggests that the mouse C2 is a non-covalently linked dimer (or possibly polymer) of subunits. The C18 and C19 embodiments are likely covalently linked dimers. Preliminary analysis suggests that the human C23 is also a covalently linked dimer (or possibly polymer) of subunits. The monomer forms exhibit reducing gel PAGE mobility corresponding to about 7.5-8 kDa.

Prediction of glycosylation sites may be made, e.g., as reported in Hansen, et al. (1995) Biochem. J. 308:801-813. However, these seem not to possess obvious sites.

Experimental determination of the N-terminus of the mature protein can be performed. N-terminal sequencing can be done on those proteins made where the terminus is not blocked.

The mouse C2B protein was expressed as an Igase/Ig fusion protein in the pCDM-8 plasmid.

#### VIII. Preparation of antibodies against CRSP

With protein produced, as above, appropriate animals are immunized to produce antibodies. Polyclonal antiserum is raised using non-purified antigen, though the resulting serum will exhibit higher background levels. Preferably, the antigen is purified using standard protein purification techniques, including, e.g., affinity chromatography using polyclonal serum indicated above. Presence of specific antibodies is detected using defined synthetic peptide fragments.

Polyclonal serum is raised against a purified antigen, purified as indicated above, or using synthetic peptides. A series of overlapping synthetic peptides which encompass all of the full length sequence, if presented to an animal, will produce serum recognizing

most linear epitopes on the protein. Such an antiserum is used to affinity purify protein, which is, in turn, used to introduce intact full length protein into another animal to produce another antiserum preparation.

- 5           Similar techniques are used to generate induce monoclonal antibodies to either unpurified antigen, or, preferably, purified antigen. Antibody fragments may also be prepared from these antibodies.

10       IX. Cellular and tissue distribution

- Distribution of the protein or gene products are determined, e.g., using immunohistochemistry with an antibody reagent, as produced above, or by screening for nucleic acids encoding the CRSP. Either hybridization or  
15       PCR methods are used to detect DNA, cDNA, or message content. Histochemistry allows determination of the specific cell types within a tissue which express higher or lower levels of message or DNA. Antibody techniques are useful to quantitate protein in a biological sample,  
20       including a liquid or tissue sample. Immunoassays are developed to quantitate protein.

- Hybridization techniques were applied. For example, adult Swiss Webster mice (Simonsen Labs, Gilroy, CA) were euthanized in a CO<sub>2</sub> atmosphere and tissues were dissected.  
25       Pregnant Swiss Webster mice were euthanized at 15 days post-coitus and embryonic tissues were dissected. Tissues were either snap frozen for RNA isolation or frozen in OCT medium (Miles, Elkhart, IN) for cryostat sectioning. Total cellular RNA was isolated from tissues by the RNazol method  
30       (Teltest, Friendswood, TX). The RNA can be used to generate cDNA libraries, thus immortalizing RNA profiles of rare or very small cell populations.

- "Reverse northern" are blots from cDNA libraries with the inserts removed, and the size determinations are  
35       based upon the size of inserts in the cDNA library, and reflect the lengths found in the cDNA library inserts, which may be less than full length where the reverse transcription was not full length. As such, size



determinations there are not reflective of the natural sizes.

The C2 embodiment was originally isolated from a Nippo infected lung tissue from a mouse. It is specifically expressed in OVA aerosol challenged mouse lung tissue, and in rag-1 mouse testis. It is expressed at lower amounts in total rag-1 thymus; bone marrow derived macrophages via GM and M-CSF; resting dendritic cells from bone marrow; and lesser amounts in rag-1 total lung, and IL-10 "knock out" (gene deletion, or KO) mesenteric lymph nodes. It is barely detectable in normal total lung; IL-10 KO colon, rag-1 total heart; highly Th2 polarized T cells, and Th2 polarized T cells. The other cDNA libraries gave no detectable signal. This suggests that the C2 is associated with lung inflammation, and in other immunologically relevant sources. Further analysis shows that the C2 is expressed in Th2 primed, ova challenged 7, 14, or 24 h; Nippo infection treated with anti-IL-5; and Aspergillus challenged, 8h; ova challenged 5x, at days 1, 2, or 3; Rag-KO mice challenged with Aspergillus after 7 or 18 h; B cell KO mice challenged with Aspergillus and anti-IL-4 after 2, 8, or 20 h; and B cell KO mice challenged with Aspergillus and anti-IL-5 after 2, 8, or 20 h. Thus, various combination compositions with other lung therapeutic entities may be suggested, e.g., with steroids, and other asthma medications.

The mouse C2b embodiment is coexpressed with C2 in the asthmatic and Nippo infected lungs, and in the Rag-1 testes. It is also expressed in certain libraries where the C2 was not detected, including large B cells from spleen, normal spleen, and Rag-1 spleen, and strongly expressed in IL-10 K.O. mouse colon and normal colon. The IL-10 KO colon is characterized by a bowel inflammatory condition.

In the cDNA libraries described above, the C2b was expressed in high amounts in IL-10 K.O. colon (X203); unstimulated large B cells from spleen (B202); total colon, normal (O212); Aspergillus challenged lung; and

total testes, rag-1 (O204). Lower amounts were expressed in aerosol challenged mouse lung tissue, Th2 primers, aerosol OVA challenge 7, 14, 23 h pooled (see Garlisi, et al. (1995) Clinical Immunology and Immunopathology 75:75-83; X206); IL-10 K.O. mesenteric lymph nodes (X203); total adult spleen, normal (O201); influenza infected lung; NZB/W spleen; IL-10 K.O. spleen (see Kuhn, et al. (1991) Cell 75:263-274; total spleen, rag-1 (O207); T cells, TH2 polarized (Mel14 bright, CD4+ cells from spleen, polarized for 7 days with IL-4 and anti-IFN- $\gamma$ ; T201); Nippostrongylus-infected lung tissue (see Coffman, et al. (1989) Science 245:308-310; X200); and Nippo infected lung treated with anti-IL-5. Detectable signals were observed in Nippo infected lung from IL-4 KO mouse; and IL-10 K.O. Peyer's patches (O202).

The C18 embodiment was highly expressed in IL-10 KO mouse colon. It was also expressed in total normal colon; rag-1 testes; and IL-10 KO mesenteric lymph nodes. It was detectable in aerosol challenged mouse lung; in B cells from LPS activated spleen; Nippo infected lung; and rat arthritic joint tissue. It was not detected in any of the other libraries tested. The observation that the mouse gene hybridizes to rat is a cross-species hybridization from mouse to rat. The other sources further suggest a role in immunological conditions, particularly inflammatory situations. Further analysis indicated that it is upregulated in IL-10 KO mice; but is downregulated, relatively, in IL-10 KO mice which have been anti-IL-12 treated.

The mouse C19 is highly expressed in rag-1 total thymus, and highly cross reacts with a rat arthritic joint. It is also present in normal rat joint; IL-10 KO colon; rag-1 total kidney; IL-10 KO mesenteric lymph nodes, rag-1 total testes; rat-1 total heart; and normal total colon. None of the other libraries exhibited a detectable signal. The distribution again suggests an immunological relevance, particularly in arthritic joint or IL-10 colon, both of which are sites of significant

inflammatory conditions. The cross hybridization with rat is also notable for this embodiment.

The expression distribution using the rat C19 gene is underway. However, various relevant samples from rat  
5 should be collected.

Distribution data for the human C10 is also being pursued, and detection in the standard human library panel has not given easily detectable signals. See panel of human libraries described in USSN 60/050,156. There  
10 is some evidence of expression in a human lung sample. Signals were detected in mouse asthmatic lung and Nippo infected lung, and in both IL-10 KO colon and normal colon. Signal was also detected in Rag-1 testes.

The C23 embodiment was originally isolated from a  
15 natural human source. It is highly expressed in the inflamed tonsil sample, and is expressed in elutriated monocytes, activated with LPS, IFNg, anti-IL-10 for 1, 2, 6, 12, 24 h pooled (M102); elutriated monocytes, activated with LPS, IFNg, IL-10 for 1, 2, 6, 12, 24 h  
20 pooled (M103); peripheral blood mononuclear cells (monocytes, T cells, NK cells, granulocytes, B cells), resting (T100); U937 premonocytic line, resting (M100); elutriated monocytes, activated with LPS, IFNg, IL-10 for 4, 16 h pooled (M107). A detectable signal is found in  
25 DC from monocytes GM-CSF, IL-4 5 days, resting (D108); DC from monocytes GM-CSF, IL-4 5 days, activated LPS 4, 16 h pooled (D109); DC from monocytes GM-CSF, IL-4 5 days, activated TNFa, monocyte supe for 4, 16 h pooled (D110). It was not detected in any of the other listed samples  
30 tested.

This distribution suggests that the C23 is important in an inflammatory response, and is found in immunologically relevant cell types.

Because of the highly specific distribution, similar  
35 tissue samples in other species will be one target for identifying other species counterparts, e.g., to primate.

## X. Chromosomal mapping

The CRSP genes can be mapped to the mouse and human chromosomes. Observations suggest that the C18 and C19 are linked. A BIOS Laboratories (New Haven, CT) mouse  
5 somatic cell hybrid panel can be combined with PCR.

Chromosomal mapping may be useful to isolate other family members where the genes are genetically linked. For example, using one of the mouse clones or the human C23 clone, chromosome walking analysis coupled with the  
10 significant homology in nucleotide sequence, will allow identification of additional new family members.

The mouse C2 has been mapped to chromosome 16, in a region which is syntenic with the part of human Chromosome 3 at which human C10 was mapped.

15

## XI. Biological assays

Mouse C2 induces the production of IgE, IgA, and IgG by purified human B cells or total human spleen cells following activation by anti-CD40 mAbs + IL-4.

20 Total human spleen cells (50,000/well) were cultured with anti-CD40 mAb 89 (10 microgram/ml) and IL-4 (400 U/ml) in the presence or absence of purified recombinant mouse C2 (100 ng/ml) in 200 microliter Yssel's medium + 10% FCS in 96 well plates (Falcon) for 14 days.

25 Alternatively, B cells (25,000/well), purified from spleen by FACS sorting following labeling with PE conjugated anti-CD20 mAbs (Leu 16; Becton-Dickensen, San Jose CA) were cultured with anti-CD40 mAb 89 (10 microgram/ml) and IL-4 (400 U/ml) in the presence or  
30 absence of purified recombinant mouse C2 (100 ng/ml) in 200 microliter Yssel's medium + 10% FCS in 96 well plates (Falcon) for 14 days. Subsequently, supernatants were harvested and the levels of human IgE, IgA, and IgG were determined by isotype specific ELISA.

35 Activation of total human spleen cells or purified human B cells with anti-CD40 and IL-4 resulted in the production of IgG, IgA, and IgE antibodies. Addition of recombinant mouse C2 to these cultures resulted in an

strong (ten fold) increase in the production of IgE, IgA, and IgG.

These results suggest that C2 may enhance the production of immunoglobulin by human B cells. This may indicate a role for C2 in allergic and inflammatory diseases, as well as in diseases in which antibody production is perturbed, such as common variable immunodeficiencies (CVI).

## 10 XII. Isolation of a Receptor

A labeled CRSP can be used as a specific binding reagent to identify its binding partner, by taking advantage of its specificity of binding, much like an antibody would be used. A binding reagent is either labeled as described above, e.g., fluorescence or otherwise, or immobilized to a substrate for panning methods. The typical chemokine receptor is a seven transmembrane receptor; and cytokine or growth hormone receptors are integral membrane proteins.

20 The binding composition, e.g., CRSP, is used to screen an expression library made from a cell line which expresses a binding partner, i.e. receptor. Standard staining techniques are used to detect or sort intracellular or surface expressed receptor, or surface expressing transformed cells are screened by panning. Screening of intracellular expression is performed by various staining or immunofluorescence procedures. See also McMahan, et al. (1991) EMBO J. 10:2821-2832.

For example, on day 0, precoat 2-chamber permanox slides with 1 ml per chamber of fibronectin, 10 ng/ml in PBS, for 30 min at room temperature. Rinse once with PBS. Then plate COS cells at  $2-3 \times 10^5$  cells per chamber in 1.5 ml of growth media. Incubate overnight at 37° C.

On day 1 for each sample, prepare 0.5 ml of a solution of 66 µg/ml DEAE-dextran, 66 µM chloroquine, and 4 µg DNA in serum free DME. For each set, a positive control is prepared, e.g., of a growth factor, cytokine, or chemokine cDNA at 1 and 1/200 dilution, and a negative mock. Rinse cells with serum free DME. Add the DNA

solution and incubate 5 hr at 37° C. Remove the medium and add 0.5 ml 10% DMSO in DME for 2.5 min. Remove and wash once with DME. Add 1.5 ml growth medium and incubate overnight.

- 5           On day 2, change the medium. On days 3 or 4, the cells are fixed and stained. Rinse the cells twice with Hank's Buffered Saline Solution (HBSS) and fix in 4% paraformaldehyde (PFA)/glucose for 5 min. Wash 3X with HBSS. The slides may be stored at -80° C after all
- 10 liquid is removed. For each chamber, 0.5 ml incubations are performed as follows. Add HBSS/saponin (0.1%) with 32 µl/ml of 1 M NaN<sub>3</sub> for 20 min. Cells are then washed with HBSS/saponin 1X. Add CRSP or CRSP/antibody complex to cells and incubate for 30 min. Wash cells twice with
- 15 HBSS/saponin. If appropriate, add first antibody for 30 min. Add second antibody, e.g., Vector anti-mouse antibody, at 1/200 dilution, and incubate for 30 min. Prepare ELISA solution, e.g., Vector Elite ABC horseradish peroxidase solution, and preincubate for 30
- 20 min. Use, e.g., 1 drop of solution A (avidin) and 1 drop solution B (biotin) per 2.5 ml HBSS/saponin. Wash cells twice with HBSS/saponin. Add ABC HRP solution and incubate for 30 min. Wash cells twice with HBSS, second wash for 2 min, which closes cells. Then add Vector
- 25 diaminobenzoic acid (DAB) for 5 to 10 min. Use 2 drops of buffer plus 4 drops DAB plus 2 drops of H<sub>2</sub>O<sub>2</sub> per 5 ml of glass distilled water. Carefully remove chamber and rinse slide in water. Air dry for a few minutes, then add 1 drop of Crystal Mount and a cover slip. Bake for 5
- 30 min at 85-90° C.

Evaluate positive staining of pools and progressively subclone to isolation of single genes responsible for the binding.

- Alternatively, CRSP reagents are used to affinity
- 35 purify or sort out cells expressing a receptor. See, e.g., Sambrook, et al. or Ausubel, et al.

Another strategy is to screen for a membrane bound receptor by panning. The receptor cDNA is constructed as described above. The ligand can be immobilized and used

to immobilize expressing cells. Immobilization may be achieved by use of appropriate antibodies which recognize, e.g., a FLAG sequence of a chemokine fusion construct, or by use of antibodies raised against the first antibodies. Recursive cycles of selection and amplification lead to enrichment of appropriate clones and eventual isolation of receptor expressing clones.

Phage expression libraries can be screened by chemokine. Appropriate label techniques, e.g., anti-FLAG antibodies, will allow specific labeling of appropriate clones.

#### XIII. Production and purification of mouse C18

A plasmid for the expression of C18 in COS (monkey kidney fibroblast) cells was constructed in the vector pCDM8 (Invitrogen, Carlsbad, CA). This plasmid was transfected into COS cells using standard techniques of electroporation. Serum-free supernatants were produced in 1 L batches by culturing  $10^8$  transfected cells for three days in Cell Factories (Nunc, Denmark).

Mouse C18 was purified from serum-free supernatants by ion-exchange chromatography on Poros Q (PerSeptive, Cambridge, MA) using a NaCl gradient in 25 mM Tris, pH 8.5. The mC18 elutes very early, around 10mM NaCl, thus achieving a substantial purification. Preparations of C18 were depleted of C18, for purposes of generating negative controls, by tumbling with the C18-specific rat monoclonal antibody 2G12 coupled to agarose beads.

#### XIV. NFS60 cell Proliferation Assay

A mouse myeloid-leukemia cell line NFS60 (see Weinstein, et al. (1986) Proc. Nat'l Acad. Sci. USA 83:5010-5014), which is a factor dependent cell line stimulatable by IL-3 or G-CSF, was assayed for responsiveness to purified C18. The C18 preparation depleted with anti-C18 antibodies was used as a control. Titration was performed to determine whether the effect was dose dependent. Preparations can also be tested for effectiveness of depletion by antibodies.

The purified mouse C18 caused (1) proliferation of the NFS60 cell line, and (2) maintained cell viability. The activity appears to be dose dependent. The effects are similar to some of the effects of G-CSF.

5       The results should be confirmed using different sources of the C18 protein. Various different methods of purification should be applied to eliminate the likelihood of artifact from contamination, e.g., with endotoxins.

10       Tests of effects with suboptimal amounts of G-CSF should be performed to determine whether there is synergistic interaction of the two factors.

Similar assays should be performed with a variety of different cells. While various cell lines should be  
15       tested, other more physiological cell types should be tested, e.g., fresh bone marrow preparations. Various fractions, e.g., progenitor cell and granulocyte enriched, should be tested to determine the extent of responsiveness. Various agar colony assays will also be  
20       tested with purified material. Hematopoietic cell responsiveness will be tested. See, e.g., Metcalf and Nicola (1995) Hemopoietic Colony-Stimulating Factors: From Biology to Clinical Applications Cambridge Univ Press.

25       With a robust proliferation assay, neutralizing antibodies may be identified.

These activities suggest that C18 probably heightens proliferation and activity of myeloid cells against infectious agents, e.g., bacteria, yeast or parasites.

30

#### XV. IL-4 inducibility of mouse C2

Mouse C2 expression is highly inducible by IL-4. IL-4 can induce C2 production by some 20-100x, even at low amounts of IL-4. As such, the C2 may serve as a  
35       sensitive means to assay for the presence of IL-4.

#### XVI. Mouse C2 genetics

Transgenic mice have been produced operably associated with a lung specific promoter. See, e.g.,



Hogan, et al. (ed.) Manipulating the Mouse Embryo: A Laboratory Manual 2d ed. CSH Press, CSH, NY. The mice are viable, but evaluation of physiological effects is proceeding.

5

#### XVII. Trafficking induced by C2

Various trafficking experiments have suggested that mouse C2 may have a chemoattracting effect on CD8 memory T cells. Thus, C2 antagonists may be useful to block chemoattraction of memory cells, e.g., in an allergy context. They may be useful in circumstances where a known exposure to antigen will occur. Alternatively, the C2 may be useful in certain tumor or other vaccination or revaccination contexts.

15 The ability of C18 to act as a leukocyte chemoattractant was examined. Supernatants of C18 gene transfected cells, or supernatants depleted of C18 using an anti-C18 Mab, were placed in the bottom well of a Costar transwell chamber and  $1 \times 10^6$  peripheral lymph node lymphocytes were placed in the top well. The apparatus was incubated at  $37^\circ \text{C}$  for 3 h to allow migration of cells from the upper to the lower chamber. A known number of beads were spiked into aliquots of the starting population and into the lower wells (to allow subsequent calculation of the number of cells present in each sample) and the cells and beads harvested and stained with fluorescent-tagged MAbs against CD4, CD8, L-selectin (CD62L), CD45RB, CD19 (B cells), and NK cells and analyzed by flow cytometry. This allows calculation of the number of cells in each subpopulation in the starting and chemoattracted population, thus allowing calculation of the chemotaxis index.

25 Preliminary results suggest that a small number of CD8+ L-selectin-low lymphocytes responded to C18 supernatants (and did not respond to depleted supes). No other subpopulation of lymphocytes were attracted. This experiment should be repeated with a broader dose range.

Confirmation would further characterize the nature of the subpopulation of CD8+-L-selectin lo cells that are

able to respond to C18. For example, it is known that the CD8+ cells located in the lung epithelium express the integrin alpha E beta 7 (that binds to E-cadherin on the epithelium), thus the expression of alpha E beta 7 on the starting and responding populations would be assessed. In addition, since in the mouse, gamma/delta T cells could constitute 1-3% of T cells in lymphoid organs, the above experiment can be repeated with an additional multicolor analysis including anti-CD3, anti- $\gamma/\delta$  TCR.

Additional studies would evaluate whether C18 is a chemoattractant for lung-derived CD4, CD8, or gamma delta T cells isolated from bronchial alveolar lavage.

All references cited herein are incorporated herein by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference in its entirety for all purposes.

Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. The specific embodiments described herein are offered by way of example only, and the invention is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

25

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Schering Corporation
- (ii) TITLE OF INVENTION: Mammalian Genes; Related Reagents
- 10 (iii) NUMBER OF SEQUENCES: 16
- 15 (iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: Schering Corporation  
(B) STREET: 2000 Galloping Hill Road  
(C) CITY: Kenilworth  
(D) STATE: New Jersey  
(E) COUNTRY: USA  
(F) ZIP: 07033
- 20 (v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: Power-Macintosh  
(C) OPERATING SYSTEM: Windows 95  
(D) SOFTWARE: MicroSoft Word
- 25 (vi) CURRENT APPLICATION DATA:  
(A) APPLICATION NUMBER:  
(B) FILING DATE: 19-JUN-1998  
(C) CLASSIFICATION:
- 30 (vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: USSN 08/878,730  
(B) FILING DATE: 19-JUN-1997
- 35 (vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: US 60/061,641  
(B) FILING DATE: 09-OCT-1997
- 40 (viii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: USSN 08/878,878  
(B) FILING DATE: 19-JUNE-1997
- 45 (viii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: Thampoe, Immac J.  
(B) REGISTRATION NUMBER: 36,322  
(C) REFERENCE/DOCKET NUMBER: DX0743X
- 50 (ix) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: (908)298-5061  
(B) TELEFAX: (908)298-5388

## (2) INFORMATION FOR SEQ ID NO:1:

- 55 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 527 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- 60 (ii) MOLECULE TYPE: cDNA
- (ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 32..364

## (ix) FEATURE:

(A) NAME/KEY: mat\_peptide

(B) LOCATION: 101..364

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

10 52 ATTCTGCCCC AGGATGCCAA CTTTGAATAG G ATG AAG ACT ACA ACT TGT TCC  
 Met Lys Thr Thr Thr Cys Ser  
 -23 -20  
 15 CTT CTC ATC TGC ATC TCC CTG CTC CAG CTG ATG GTC CCA GTG AAT ACT  
 100  
 Leu Leu Ile Cys Ile Ser Leu Leu Gln Leu Met Val Pro Val Asn Thr  
 -15 -10 -5  
 20 GAT GAG ACC ATA GAG ATT ATC GTG GAG AAT AAG GTC AAG GAA CTT CTT  
 148  
 Asp Glu Thr Ile Glu Ile Ile Val Glu Asn Lys Val Lys Glu Leu Leu  
 1 5 10 15  
 25 GCC AAT CCA GCT AAC TAT CCC TCC ACT GTA ACG AAG ACT CTC TCT TGC  
 196  
 Ala Asn Pro Ala Asn Tyr Pro Ser Thr Val Thr Lys Thr Leu Ser Cys  
 20 25 30  
 30 ACT AGT GTC AAG ACT ATG AAC AGA TGG GCC TCC TGC CCT GCT GGG ATG  
 244  
 Thr Ser Val Lys Thr Met Asn Arg Trp Ala Ser Cys Pro Ala Gly Met  
 35 40 45  
 35 ACT GCT ACT GGG TGT GCT TGT GGC TTT GCC TGT GGA TCT TGG GAG ATC  
 292  
 Thr Ala Thr Gly Cys Ala Cys Gly Phe Ala Cys Gly Ser Trp Glu Ile  
 50 55 60  
 40 CAG AGT GGA GAT ACT TGC AAC TGC CTG TGC TTA CTC GTT GAC TGG ACC  
 340  
 Gln Ser Gly Asp Thr Cys Asn Cys Leu Cys Leu Leu Val Asp Trp Thr  
 65 70 75 80  
 45 ACT GCC CGC TGC TGC CAA CTG TCC TAAGAATGAA GAGGTGGAGA ACCCAGCTTT  
 394  
 Thr Ala Arg Cys Cys Gln Leu Ser  
 85  
 50 GATATGATGA ATCTAACAAA AACTGCAGTC TCAATTTGGA AATCTGACTC ATGTGCCTTT  
 454  
 AAATGTGTTC ATATTGCCCA TTTACCCTGC TTCTTGAAAT GCTTCTTGAA AAATAAAGAC  
 514  
 55 AAATTTGCAT GTG  
 527

## (2) INFORMATION FOR SEQ ID NO:2:

60

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 111 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

5 Met Lys Thr Thr Thr Cys Ser Leu Leu Ile Cys Ile Ser Leu Leu Gln  
-23 -20 -15 -10

10 Leu Met Val Pro Val Asn Thr Asp Glu Thr Ile Glu Ile Ile Val Glu  
-5 1 5

Asn Lys Val Lys Glu Leu Leu Ala Asn Pro Ala Asn Tyr Pro Ser Thr  
10 15 20 25

15 Val Thr Lys Thr Leu Ser Cys Thr Ser Val Lys Thr Met Asn Arg Trp  
30 35 40

Ala Ser Cys Pro Ala Gly Met Thr Ala Thr Gly Cys Ala Cys Gly Phe  
45 50 55

20 Ala Cys Gly Ser Trp Glu Ile Gln Ser Gly Asp Thr Cys Asn Cys Leu  
60 65 70

25 Cys Leu Leu Val Asp Trp Thr Thr Ala Arg Cys Cys Gln Leu Ser  
75 80 85

(2) INFORMATION FOR SEQ ID NO:3:

30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 574 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: cDNA

40 (ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 70..402

45 (ix) FEATURE:  
(A) NAME/KEY: mat\_peptide  
(B) LOCATION: 139..402

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

50 AGCATCTCAT CTGGCCAGGT CCTGGAACCT TTCCTGAGAT TCTGCCCTAG GATGCTGACT  
60

TTCAACAAG ATG AAG ACT ACA ACT TGT TCC CTT CTC ATC TGC ATC TCC  
108

55 Met Lys Thr Thr Thr Cys Ser Leu Leu Ile Cys Ile Ser  
-23 -20 -15

CTT CTC CAG CTG ATG GTC CCA GTG AAT ACT GAG GGG ACC TTA GAA TCT  
156

60 Leu Leu Gln Leu Met Val Pro Val Asn Thr Glu Gly Thr Leu Glu Ser  
-10 -5 1 5

ATT GTG GAG AAA AAG GTC AAG GAA CTT CTT GCC AAT CGA GAT GAC TGT  
204

```

      Ile Val Glu Lys Lys Val Lys Glu Leu Leu Ala Asn Arg Asp Asp Cys
                10                15                20

5    CCC TCC ACT GTA ACA AAG ACT TTC TCC TGT ACT AGT ATC ACG GCT TCA
    252
      Pro Ser Thr Val Thr Lys Thr Phe Ser Cys Thr Ser Ile Thr Ala Ser
                25                30                35
      GGC AGA CTG GCC TCC TGT CCT TCT GGA ATG ACT GTC ACT GGT TGT GCT
    300
10   Gly Arg Leu Ala Ser Cys Pro Ser Gly Met Thr Val Thr Gly Cys Ala
      40                45                50

      TGT GGC TAT GGC TGT GGA TCT TGG GAT ATC CGG GAT GGA AAT ACT TGC
    348
15   Cys Gly Tyr Gly Cys Gly Ser Trp Asp Ile Arg Asp Gly Asn Thr Cys
      55                60                65                70

      CAC TGT CAG TGC TCA ACA ATG GAC TGG GCC ACC GCC CGT TGC TGC CAA
    396
20   His Cys Gln Cys Ser Thr Met Asp Trp Ala Thr Ala Arg Cys Cys Gln
      75                80                85

      CTG GCC TAAGAATGAG GAAGCTGAGA ACCTAGCTTT GAAATGAAGA CTATAACAAA
    452
25   Leu Ala

      AGCACAATCC CAACTTGGAA ACCTGGCTCA TATCCCATTG ATGAATTCAT ATTGTCCATT
    512
30   AGCCCTGCTT CTTGAAAAAA ATAAAGACAA ATTTGCACGT GTCTGTAAAA AAAAA-AAAAA
    572

      AA
35   574

```

## (2) INFORMATION FOR SEQ ID NO:4:

```

40   (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 111 amino acids
      (B) TYPE: amino acid
      (D) TOPOLOGY: linear

45   (ii) MOLECULE TYPE: protein

      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

50   Met Lys Thr Thr Thr Cys Ser Leu Leu Ile Cys Ile Ser Leu Leu Gln
      -23      -20      -15      -10

      Leu Met Val Pro Val Asn Thr Glu Gly Thr Leu Glu Ser Ile Val Glu
            -5                1                5

55   Lys Lys Val Lys Glu Leu Leu Ala Asn Arg Asp Asp Cys Pro Ser Thr
      10                15                20                25

      Val Thr Lys Thr Phe Ser Cys Thr Ser Ile Thr Ala Ser Gly Arg Leu
            30                35                40

60   Ala Ser Cys Pro Ser Gly Met Thr Val Thr Gly Cys Ala Cys Gly Tyr
            45                50                55

      Gly Cys Gly Ser Trp Asp Ile Arg Asp Gly Asn Thr Cys His Cys Gln

```

97

70 75 80

TGC TGC CGA ATG GCT TAAGAATGAG GAGGTTGAGA AACCAATTTT AAAATGATGA  
457

5 Cys Cys Arg Met Ala  
85

CCATAATGAA ACCACGGTCT CGACCAGGAA ACCTGACTCA TTGTCTTCAT ATTACTAAAT  
517

10 AATTCTTCTT GAATAATAAA GGCAGACCTG TACCTTT  
554

15 (2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 105 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

25 Met Lys Pro Thr Leu Cys Phe Leu Phe Ile Leu Val Ser Leu Phe Pro  
-19 -15 -10 -5

30 Leu Ile Val Pro Gly Asn Ala Gln Cys Ser Phe Glu Ser Leu Val Asp  
1 5 10

Gln Arg Ile Lys Glu Ala Leu Ser Arg Gln Glu Pro Lys Thr Ile Ser  
15 20 25

35 Cys Thr Ser Val Thr Ser Ser Gly Arg Leu Ala Ser Cys Pro Ala Gly  
30 35 40 45

Met Val Val Thr Gly Cys Ala Cys Gly Tyr Gly Cys Gly Ser Trp Asp  
50 55 60

40 Ile Arg Asn Gly Asn Thr Cys His Cys Gln Cys Ser Val Met Asp Trp  
65 70 75

Ala Ser Ala Arg Cys Cys Arg Met Ala  
80 85

45

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 560 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

50

55 (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:  
(A) NAME/KEY: CDS  
(B) LOCATION: 64..405

60

(ix) FEATURE:  
(A) NAME/KEY: mat\_peptide  
(B) LOCATION: 124..405



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

5 GACAGGAGCT AATACCCAGA ACTGAGTTGT GTCCTGCTAA GTCCTCTGCC ACGTACCCAC  
 60  
 GGG ATG AAG AAC CTT TCA TTT CCC CTC CTT TTC CTT TTC CTT GTC  
 108  
 10 Met Lys Asn Leu Ser Phe Pro Leu Leu Phe Leu Phe Phe Leu Val  
 -20 -15 -10  
 CCT GAA CTG CTG GGC TCC AGC ATG CCA CTG TGT CCC ATC GAT GAA GCC  
 156  
 15 Pro Glu Leu Leu Gly Ser Ser Met Pro Leu Cys Pro Ile Asp Glu Ala  
 -5 1 5 10  
 ATC GAC AAG AAG ATC AAA CAA GAC TTC AAC TCC CTG TTT CCA AAT GCA  
 204  
 20 Ile Asp Lys Lys Ile Lys Gln Asp Phe Asn Ser Leu Phe Pro Asn Ala  
 15 20 25  
 ATA AAG AAC ATT GGC TTA AAT TGC TGG ACA GTC TCC TCC AGA GGG AAG  
 252  
 25 Ile Lys Asn Ile Gly Leu Asn Cys Trp Thr Val Ser Ser Arg Gly Lys  
 30 35 40  
 TTG GCC TCC TGC CCA GAA GGC ACA GCA GTC TTG AGC TGC TCC TGT GGC  
 300  
 30 Leu Ala Ser Cys Pro Glu Gly Thr Ala Val Leu Ser Cys Ser Cys Gly  
 45 50 55  
 TCT GCC TGT GGC TCG TGG GAC ATT CGT GAA GAA AAA GTG TGT CAC TGC  
 348  
 35 Ser Ala Cys Gly Ser Trp Asp Ile Arg Glu Glu Lys Val Cys His Cys  
 60 65 70 75  
 CAG TGT GCA AGG ATA GAC TGG ACA GCA GCC CGC TGC TGT AAG CTG CAG  
 396  
 40 Gln Cys Ala Arg Ile Asp Trp Thr Ala Ala Arg Cys Cys Lys Leu Gln  
 80 85 90  
 GTC GCT TCC TGATGTCGGG GAAGTGAGCG TGTTTCCAG CACAGCCACC  
 445  
 45 Val Ala Ser  
 CGTTCCTGTA GCTCCAGAGA TGTCTGATGT CCTCCGGTCT CTACAGGCAC CTGCACTCAC  
 505  
 50 GTGCGCGAAT CCACACACAA GCACACATAC TTAAAAATAA AACAAAACAG GCTGG  
 560

## 55 (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- 60 (A) LENGTH: 114 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Lys Asn Leu Ser Phe Pro Leu Leu Phe Leu Phe Phe Leu Val Pro  
 -20 -15 -10 -5

5 Glu Leu Leu Gly Ser Ser Met Pro Leu Cys Pro Ile Asp Glu Ala Ile  
 1 5 10

Asp Lys Lys Ile Lys Gln Asp Phe Asn Ser Leu Phe Pro Asn Ala Ile  
 15 20 25

10 Lys Asn Ile Gly Leu Asn Cys Trp Thr Val Ser Ser Arg Gly Lys Leu  
 30 35 40

Ala Ser Cys Pro Glu Gly Thr Ala Val Leu Ser Cys Ser Cys Gly Ser  
 15 45 50 55 60

Ala Cys Gly Ser Trp Asp Ile Arg Glu Glu Lys Val Cys His Cys Gln  
 65 70 75

20 Cys Ala Arg Ile Asp Trp Thr Ala Ala Arg Cys Cys Lys Leu Gln Val  
 80 85 90

Ala Ser

25 (2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 572 base pairs  
 30 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA  
 35

(ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 29..370  
 40

(ix) FEATURE:  
 (A) NAME/KEY: mat\_peptide  
 (B) LOCATION: 80..370  
 45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CTGAGCTCTC TGCCACGTAC TTAACAGG ATG AAG AAC CTT TCA TTT CTC CTC  
 50 52

Met Lys Asn Leu Ser Phe Leu Leu  
 -17 -15 -10

CTT TTC CTT TTC TTC CTT GTC CTG GGG CTG CTG GGC CCC AGC ATG TCA  
 55 100

Leu Phe Leu Phe Phe Leu Val Leu Gly Leu Leu Gly Pro Ser Met Ser  
 -5 1 5

CTG TGT CCC ATG GAT GAA GCC ATC AGC AAG AAG ATC AAT CAA GAC TTC  
 60 148

Leu Cys Pro Met Asp Glu Ala Ile Ser Lys Lys Ile Asn Gln Asp Phe  
 10 15 20

AGC TCC CTA CTG CCA GCT GCA ATG AAG AAC ACT GTC CTA CAT TGC TGG  
 196  
 Ser Ser Leu Leu Pro Ala Ala Met Lys Asn Thr Val Leu His Cys Trp  
 25 30 35  
 5  
 TCA GTC TCC TCC AGA GGG AGG CTG GCC TCC TGC CCA GAA GGC ACA ACC  
 244  
 Ser Val Ser Ser Arg Gly Arg Leu Ala Ser Cys Pro Glu Gly Thr Thr  
 40 45 50 55  
 10  
 GTC ACT AGC TGC TCC TGT GGC TCT GGC TGT GGC TCA TGG GAC GTC CGT  
 292  
 Val Thr Ser Cys Ser Cys Gly Ser Gly Cys Gly Ser Trp Asp Val Arg  
 60 65 70  
 15  
 GAG GAT ACA ATG TGT CAC TGC CAG TGC GGA AGC ATA GAC TGG ACA GCG  
 340  
 Glu Asp Thr Met Cys His Cys Gln Cys Gly Ser Ile Asp Trp Thr Ala  
 75 80 85  
 20  
 GCC CGC TGC TGT ACC CTG CGG GTT GGT TCC TGAGGACGGT TGATTGAGAA  
 390  
 Ala Arg Cys Cys Thr Leu Arg Val Gly Ser  
 90 95  
 25  
 CTGAGCTTGC CCTCCGAGTG CTGCCGAGGG ATGAGCTTGC CCACCATGCC CTGCAGAGGA  
 450  
 GGGATGGGGA TGGGGAGAGC GCAGGGGGCA GGAAACGAGA TGAGGGTTTG GAAATACACA  
 510  
 30  
 ATGGGATGAT GGTGGTGATA AAGATGCACG GTAAAGTGA AAAAAAAAAA AAAAAAAAAA  
 570  
 35  
 AA  
 572

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 114 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

50 Met Lys Asn Leu Ser Phe Leu Leu Leu Phe Leu Phe Phe Leu Val Leu  
 -17 -15 -10 -5  
 Gly Leu Leu Gly Pro Ser Met Ser Leu Cys Pro Met Asp Glu Ala Ile  
 1 5 10 15  
 55 Ser Lys Lys Ile Asn Gln Asp Phe Ser Ser Leu Leu Pro Ala Ala Met  
 20 25 30  
 60 Lys Asn Thr Val Leu His Cys Trp Ser Val Ser Ser Arg Gly Arg Leu  
 35 40 45  
 Ala Ser Cys Pro Glu Gly Thr Thr Val Thr Ser Cys Ser Cys Gly Ser  
 50 55 60

Gly Cys Gly Ser Trp Asp Val Arg Glu Asp Thr Met Cys His Cys Gln  
 65 70 75  
 Cys Gly Ser Ile Asp Trp Thr Ala Ala Arg Cys Cys Thr Leu Arg Val  
 5 80 85 90 95  
 Gly Ser

10 (2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 603 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20 (ix) FEATURE:

(A) NAME/KEY: CDS  
 (B) LOCATION: 108..440

25 (ix) FEATURE:

(A) NAME/KEY: mat\_peptide  
 (B) LOCATION: 168..440

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGCACGAGGC CACGTTGTCT TCTTTCCTTC ACCACCACCC AGGAGCTCAG AGATCTAAGC  
 60  
 35 TGCTTTCCAT CTTTCTCCC AGCCCCAGGA CACTGACTCT GTACAGG ATG GGG CCG  
 116  
 Met Gly Pro  
 -20  
 40 TCC TCT TGC CTC CTT CTC ATC CTA ATC CCC CTT CTC CAG CTG ATC AAC  
 164  
 Ser Ser Cys Leu Leu Leu Ile Leu Ile Pro Leu Leu Gln Leu Ile Asn  
 -15 -10 -5  
 45 CCG GGG AGT ACT CAG TGT TCC TTA GAC TCC GTT ATG GAT AAG AAG ATC  
 212  
 Pro Gly Ser Thr Gln Cys Ser Leu Asp Ser Val Met Asp Lys Lys Ile  
 1 5 10 15  
 50 AAG GAT GTT CTC AAC AGT CTA GAG TAC AGT CCC TCT CCT ATA AGC AAG  
 260  
 Lys Asp Val Leu Asn Ser Leu Glu Tyr Ser Pro Ser Pro Ile Ser Lys  
 20 25 30  
 55 AAG CTC TCG TGT GCT AGT GTC AAA AGC CAA GGC AGA CCG TCC TCC TGC  
 308  
 Lys Leu Ser Cys Ala Ser Val Lys Ser Gln Gly Arg Pro Ser Ser Cys  
 35 40 45  
 60 CCT GCT GGG ATG GCT GTC ACT GGC TGT GCT TGT GGC TAT GGC TGT GGT  
 356  
 Pro Ala Gly Met Ala Val Thr Gly Cys Ala Cys Gly Tyr Gly Cys Gly  
 50 55 60

TCG TGG GAT GTT CAG CTG GAA ACC ACC TGC CAC TGC CAG TGC AGT GTG  
 404  
 Ser Trp Asp Val Gln Leu Glu Thr Thr Cys His Cys Gln Cys Ser Val  
     65                    70                    75

5 GTG GAC TGG ACC ACT GCC CGC TGC TGC CAC CTG ACC TGACAGGGAG  
 450  
 Val Asp Trp Thr Thr Ala Arg Cys Cys His Leu Thr  
     80                    85                    90

10 GAGGCTGAGA ACTCAGTTTT GTGACCATGA CAGTAATGAA ACCAGGGTCC CAACCAAGAA  
 510

15 ATCTAACTCA AACGTCCCAC TTCATTTGTT CCATTCCTGA TTCTTGGGTA ATAAAGACAA  
 570

ACTTTGTACC TCAAAAAAAAA AAAAAAAAAAAA AAA  
 603

20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 111 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: protein

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Gly Pro Ser Ser Cys Leu Leu Leu Ile Leu Ile Pro Leu Leu Gln  
 -20                    -15                    -10                    -5

35 Leu Ile Asn Pro Gly Ser Thr Gln Cys Ser Leu Asp Ser Val Met Asp  
                     1                    5                    10

Lys Lys Ile Lys Asp Val Leu Asn Ser Leu Glu Tyr Ser Pro Ser Pro  
                     15                    20                    25

40 Ile Ser Lys Lys Leu Ser Cys Ala Ser Val Lys Ser Gln Gly Arg Pro  
                     30                    35                    40

Ser Ser Cys Pro Ala Gly Met Ala Val Thr Gly Cys Ala Cys Gly Tyr  
 45                    45                    50                    55                    60

Gly Cys Gly Ser Trp Asp Val Gln Leu Glu Thr Thr Cys His Cys Gln  
                     65                    70                    75

50 Cys Ser Val Val Asp Trp Thr Thr Ala Arg Cys Cys His Leu Thr  
                     80                    85                    90

55

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 453 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

60

(ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 47..370

5

## (ix) FEATURE:

(A) NAME/KEY: mat\_peptide

(B) LOCATION: 101..370

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTGTGCCGGA TTTGGTTAGC TGAGCCCACC GAGAGGCGCC TGCAGG ATG AAA GCT  
 55  
 15 Met Lys Ala  
 -18  
 CTC TGT CTC CTC CTC CTC CCT GTC CTG GGG CTG TTG GTG TCT AGC AAG  
 103  
 20 Leu Cys Leu Leu Leu Leu Pro Val Leu Gly Leu Leu Val Ser Ser Lys  
 -15 -10 -5 1  
 ACC CTG TGC TCC ATG GAA GAA GCC ATC AAT GAG AGG ATC CAG GAG GTC  
 151  
 25 Thr Leu Cys Ser Met Glu Glu Ala Ile Asn Glu Arg Ile Gln Glu Val  
 5 10 15  
 GCC GGC TCC CTA ATA TTT AGG GCA ATA AGC AGC ATT GGC CTG GAG TGC  
 199  
 30 Ala Gly Ser Leu Ile Phe Arg Ala Ile Ser Ser Ile Gly Leu Glu Cys  
 20 25 30  
 CAG AGC GTC ACC TCC AGG GGG GAC CTG GCT ACT TGC CCC CGA GGC TTC  
 247  
 35 Gln Ser Val Thr Ser Arg Gly Asp Leu Ala Thr Cys Pro Arg Gly Phe  
 35 40 45  
 GCC GTC ACC GGC TGC ACT TGT GGC TCC GCC TGT GGC TCG TGG GAT GTG  
 295  
 40 Ala Val Thr Gly Cys Thr Cys Gly Ser Ala Cys Gly Ser Trp Asp Val  
 50 55 60 65  
 CGC GCC GAG ACC ACA TGT CAC TGC CAG TGC GCG GGC ATG GAC TGG ACC  
 343  
 45 Arg Ala Glu Thr Thr Cys His Cys Gln Cys Ala Gly Met Asp Trp Thr  
 70 75 80  
 GGA GCG CGC TGC TGT CGT GTG CAG CCC TGAGGTCGCG CGCAGCGCGT  
 390  
 50 Gly Ala Arg Cys Cys Arg Val Gln Pro  
 85 90  
 GCACAGCGCG GCGGAGGCG GCTCCAGGTC CGGAGGGGTT GCGGGGAGC TGGAAATAAA  
 450  
 55 CCT  
 453

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: Cleavage-site  
 (B) LOCATION: 16..17

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```

Met Lys Ala Leu Cys Leu Leu Leu Leu Pro Val Leu Gly Leu Leu Val
1      5      10      15
Ser Ser Lys Thr Leu Cys Ser Met Glu Glu Ala Ile Asn Glu Arg Ile
20      25      30
Gln Glu Val Ala Gly Ser Leu Ile Phe Arg Ala Ile Ser Ser Ile Gly
35      40      45
Leu Glu Cys Gln Ser Val Thr Ser Arg Gly Asp Leu Ala Thr Cys Pro
50      55      60
Arg Gly Phe Ala Val Thr Gly Cys Thr Cys Gly Ser Ala Cys Gly Ser
65      70      75      80
Trp Asp Val Arg Ala Glu Thr Thr Cys His Cys Gln Cys Ala Gly Met
85      90      95
Asp Trp Thr Gly Ala Arg Cys Cys Arg Val Gln Pro
100     105

```

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```

TGTGGCTHYG SCTGTGGMTC KTGG
24

```

## (2) INFORMATION FOR SEQ ID NO:16:

5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: cDNA

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CAGCAGCGSG CWSHKGTCCTCA GTC  
23



## WHAT IS CLAIMED IS:

1. A composition of matter selected from the group consisting of:

- 5 a) a substantially pure or recombinant C2 polypeptide exhibiting identity over a length of at least 12 contiguous amino acids to SEQ ID NO: 2;
- " b) a natural sequence C2 of SEQ ID NO: 2;
- " c) a fusion protein comprising C2 sequence;
- 10 d) a substantially pure or recombinant C2b polypeptide exhibiting identity over a length of at least 12 contiguous amino acids to SEQ ID NO: 4;
- e) a natural sequence C2b of SEQ ID NO: 4;
- 15 f) a fusion protein comprising C2b sequence;
- g) a substantially pure or recombinant C18 polypeptide exhibiting identity over a length of at least 12 contiguous amino acids to SEQ ID NO: 6;
- 20 h) a natural sequence C18 of SEQ ID NO: 6;
- i) a fusion protein comprising C18 sequence;
- j) a substantially pure or recombinant C19 polypeptide exhibiting identity over a length of at least 12 contiguous amino acids to SEQ ID ,
- 25 NO: 8 or 10;
- k) a natural sequence C19 of SEQ ID NO: 8 or 10;
- l) a fusion protein comprising C19 sequence;
- m) a substantially pure or recombinant C10 polypeptide exhibiting identity over a length
- 30 of at least 12 contiguous amino acids to SEQ ID NO: 12;
- n) a natural sequence C10 of SEQ ID NO: 12;
- o) a fusion protein comprising C10 sequence;
- p) a substantially pure or recombinant C23 protein
- 35 or peptide exhibiting at least about 85% sequence identity over a length of at least about 12 amino acids to SEQ ID NO: 2;
- q) a natural sequence C23 of SEQ ID NO: 2; and
- r) a fusion protein comprising C23 sequence.

2. A substantially pure or isolated polypeptide comprising a segment exhibiting sequence identity to a corresponding portion of a C2, C2b, C18, C19, C10 or C23 of Claim 1, wherein:
- a) said identity is over at least 15 contiguous amino acids;
  - b) said identity is over at least 19 contiguous amino acids; or
  - 10 c) said identity is over at least 25 contiguous amino acids.
3. The composition of matter of Claim 1, wherein said:
- 15 a) C2 or C2b comprises a mature sequence of Table 1;
  - b) C18 comprises a mature sequence of Table 2;
  - c) C19 comprises a mature sequence of Table 3;
  - d) C10 comprises a mature sequence of Table 4;
  - 20 e) C23 comprises a mature sequence of Table 6;
  - f) polypeptide:
    - i) is from a warm blooded animal selected from a mammal, including a rodent or primate;
    - ii) comprises at least 27 contiguous amino acids of SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14;
    - 25 iii) exhibits a plurality of said lengths exhibiting said identity;
    - iv) is a natural allelic variant of SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14
    - 30 v) has a length at least about 30 amino acids;
    - vi) exhibits at least two non-overlapping epitopes which are specific for a mammalian C2, C2b, C18, C19, C10 or C23;
    - 35 vii) exhibits a sequence identity over at least 33 amino acids to SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14;

- viii) exhibits at least two non-overlapping epitopes which are specific for SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14;
- ix) exhibits sequence identity over a length of at least about 20 amino acids to SEQ ID NO: 2, 4, 6, 8, 10, 12 or 14;
- x) is not glycosylated;
- xi) has a molecular weight of at least 3 kD;
- xii) is a synthetic polypeptide;
- xiii) is attached to a solid substrate;
- xiv) is conjugated to another chemical moiety;
- xv) is a 5-fold or less substitution from natural sequence; or
- xvi) is a deletion or insertion variant from a natural sequence.

4. A composition comprising:

- a) a sterile C2 or C2b polypeptide of Claim 1,
- b) said C2 or C2b polypeptide of Claim 1 and a carrier, wherein said carrier is:
- i) an aqueous compound, including water, saline, and/or buffer; and/or
- ii) formulated for oral, rectal, nasal, topical, or parenteral administration;
- c) a sterile C18 polypeptide of Claim 1;
- d) said C18 polypeptide of Claim 1 and a carrier, wherein said carrier is:
- i) an aqueous compound, including water, saline, and/or buffer; and/or
- ii) formulated for oral, rectal, nasal, topical, or parenteral administration;
- e) a sterile C19 polypeptide of Claim 1;
- f) said C19 polypeptide of Claim 1 and a carrier, wherein said carrier is:
- i) an aqueous compound, including water, saline, and/or buffer; and/or
- ii) formulated for oral, rectal, nasal, topical, or parenteral administration
- g) a sterile C10 polypeptide of Claim 1;

- h) said C10 polypeptide of Claim 1 and a carrier,  
wherein said carrier is:
- i) an aqueous compound, including water,  
saline, and/or buffer; and/or
  - 5 ii) formulated for oral, rectal, nasal,  
topical, or parenteral administration.
- i) a sterile C23 polypeptide of Claim 1; or
- j) said C23 polypeptide of Claim 1 and a carrier,  
wherein said carrier is:
- 10 i) an aqueous compound, including water,  
saline, and/or buffer; and/or
  - ii) formulated for oral, rectal, nasal,  
topical, or parenteral administration
- 15 5. The fusion protein of Claim 1, comprising:
- a) mature protein sequence of Table 1, 2, 3, 4, or  
6;
  - b) a detection or purification tag, including a  
FLAG, His6, or Ig sequence; or
  - 20 c) sequence of another cytokine or growth factor  
protein.
6. A kit comprising a polypeptide of Claim 1,  
and:
- 25 a) a compartment comprising said polypeptide;  
and/or
  - b) instructions for use or disposal of reagents in  
said kit.
- 30 7. A binding compound comprising an antigen  
binding portion from an antibody, which specifically  
binds to a natural C2, C18, C19, C10 or C23 polypeptide  
of Claim 1, wherein:
- a) said polypeptide is a rodent C2, C2b, C18, or  
35 C19 protein;
  - b) said polypeptide is a primate C10 or C23  
protein;
  - c) said binding compound is an Fv, Fab, or Fab2  
fragment;

- d) said binding compound is conjugated to another chemical moiety; or
- e) said antibody:
- 5 i) is raised against a peptide sequence of a mature polypeptide of Table 1, 2, 3, 4 or 6;
  - ii) is raised against a mature C2, C2b, C18, C19, C10 or C23;
  - 10 iii) is raised to a purified C2, C2b, C18, C19, C10 or C23;
  - iv) is immunoselected;
  - v) is a polyclonal antibody;
  - vi) binds to a denatured C2, C2b, C18, C19, C10 or C23;
  - 15 vii) exhibits a  $K_d$  to antigen of at least 30  $\mu M$ ;
  - viii) is attached to a solid substrate, including a bead or plastic membrane;
  - ix) is in a sterile composition; or
  - 20 x) is detectably labeled, including a radioactive or fluorescent label.

8. A kit comprising said binding compound of Claim 7, and:
- 25 a) a compartment comprising said binding compound; and/or
  - b) instructions for use or disposal of reagents in said kit.

- 30 9. A method of:
- A) making an antibody of Claim 7, comprising immunizing an immune system with an immunogenic amount of:
- 35 a) a rodent C2 polypeptide;
  - b) a rodent C2b polypeptide;
  - c) a rodent C18 polypeptide;
  - d) a rodent C19 polypeptide;
  - e) a primate C10 polypeptide; or
  - f) a primate C23 polypeptide;

- thereby causing said antibody to be produced;  
or
- B) producing an antigen:antibody complex,  
comprising contacting:
- 5 a) a rodent C2 polypeptide with an antibody of  
Claim 7;
- b) a rodent C2b polypeptide with an antibody  
of Claim 7; or
- 10 c) a rodent C18 polypeptide with an antibody  
of Claim 7;
- d) a rodent C19 polypeptide with an antibody  
of Claim 7;
- e) a primate C10 polypeptide with an antibody  
of Claim 7; or
- 15 f) a primate C23 polypeptide with an antibody  
of Claim 7;
- thereby allowing said complex to form.
10. A composition comprising:
- 20 a) a sterile binding compound of Claim 7, or
- b) said binding compound of Claim 7 and a carrier,  
wherein said carrier is:
- i) an aqueous compound, including water,  
saline, and/or buffer; and/or
- 25 ii) formulated for oral, rectal, nasal,  
topical, or parenteral administration.
11. An isolated or recombinant nucleic acid  
encoding a polypeptide or fusion protein of Claim 1,  
30 wherein:
- a) said C family protein is from a mammal,  
including a rodent or primate; or
- b) said nucleic acid:
- i) encodes an antigenic peptide sequence of  
35 Table 1, 2, 3, 4 or 6;
- ii) encodes a plurality of antigenic peptide  
sequences of Table 1, 2, 3, 4 or 6;
- iii) exhibits at least about 80% identity to a  
natural cDNA encoding said segment;

- iv) is an expression vector;
- v) further comprises an origin of replication;
- vi) is from a natural source;
- vii) comprises a detectable label;
- 5 viii) comprises synthetic nucleotide sequence;
- ix) is less than 6 kb, preferably less than 3 kb;
- x) is from a mammal, including a rodent or primate;
- 10 xi) comprises a natural full length coding sequence;
- xii) is a hybridization probe for a gene encoding said C family protein; or
- 15 xiii) is a PCR primer, PCR product, or mutagenesis primer.

12. A cell or tissue comprising a recombinant nucleic acid of Claim 11.

20 13. The cell of Claim 12, wherein said cell is:

- a) a prokaryotic cell;
- b) a eukaryotic cell;
- c) a bacterial cell;
- d) a yeast cell;
- 25 e) an insect cell;
- f) a mammalian cell;
- g) a mouse cell;
- h) a primate cell; or
- i) a human cell.

30

14. A kit comprising said nucleic acid of Claim 11, and:

- a) a compartment comprising said nucleic acid;
- b) a compartment further comprising a C2, C2b, C18, C19, C10 or C23 polypeptide; and/or
- 35 c) instructions for use or disposal of reagents in said kit.

15. A method of:

- A) making a polypeptide, comprising expressing said nucleic acid of Claim 11, thereby producing said polypeptide; or
- 5 B) making a duplex nucleic acid, comprising contacting said nucleic acid of Claim 11 with a hybridizing nucleic acid, thereby allowing said duplex to form.

16. A nucleic acid which:
- 10 a) hybridizes under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 1;
- b) hybridizes under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 3;
- 15 c) hybridizes under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 5;
- d) hybridizes under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 7;
- e) hybridizes under wash conditions of 30° C and less than 2 M salt to SEQ ID NO: 9;
- 20 f) hybridizes under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 11;
- g) hybridizes under wash conditions of 30° C and less than 2M salt to SEQ ID NO: 13;
- 25 h) exhibits at least about 85% identity over a stretch of at least about 30 nucleotides to a rodent C2 or C2b;
- i) exhibits at least about 85% identity over a stretch of at least about 30 nucleotides to a rodent C18;
- 30 j) exhibits at least about 85% identity over a stretch of at least about 30 nucleotides to a rodent C19;
- k) exhibits at least about 85% identity over a stretch of at least about 30 nucleotides to a primate C10; or
- 35 l) exhibits at least about 85% identity over a stretch of at least about 30 nucleotides to a primate C23



17. The nucleic acid of Claim 16, wherein:

a) said wash conditions are:

i) at 45° C and/or 500 mM salt; or

ii) at 55° C and/or 150 mM salt; or

5 b) said identity is:

i) at least 90% and/or said stretch is at  
least 55 nucleotides; or

ii) at least 95% and/or said stretch is at  
least 75 nucleotides.

10

18. A method of modulating physiology or  
development of a cell or tissue culture cells comprising  
contacting said cell with an agonist or antagonist of a  
C2, C2b, C18, C19, C10 or C23.

15

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/12236

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C07K14/47 C07K16/18 C12Q1/68 G01N33/577

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	M. MARRA ET AL.: "The WashU-HHMI Mouse EST Project" EMBL SEQUENCE DATABASE, 29 March 1997, XP002078354 Heidelberg, FRG vb70e06.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA clone 762370 5'; Accession no. AA273994; ---	11-13, 17
X	M. MARRA ET AL.: "The WashU-HHMI EST Project" EMBL SEQUENCE DATABASE, 12 March 1997, XP002078355 Heidelberg, FRG my52f08.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA clone 699495 5'; Accession no. AA245405 5'; --- -/--	11-13, 17

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

### \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

23 September 1998

Date of mailing of the international search report

01/10/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2230 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Hornig, H

## INTERNATIONAL SEARCH REPORT

Intel. Application No

PCT/US 98/12236

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	M. MARRA ET AL.: "The WashU-HHMI Mouse EST Project" EMBL SEQUENCE DATABASE, 20 May 1997, XP002078356 Heidelberg, FRG ve76b03.r1 Soares mouse mammary gland NbMMG Mus. musculus cDNA clone 832109 5'; Accession no. AA423511; ---	11-13, 17
X	M. MARRA ET AL.: "The WashU-HHMI Mouse EST Project" EMBL SEQUENCE DATABASE, 14 June 1997, XP002078357 Heidelberg, FRG ve01d11.r1 Soares mouse NbMH Mus musculus cDNA clone 808821 5'; Accession no. AA467519; ---	11-13, 17
X	M.D. ADAMS ET AL.: "Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequences" EMBL SEQUENCE DATABASE, 18 April 1997, XP002078358 Heidelberg, FRG & NATURE, vol. 377, 1995, pages 3-174, MACMILLAN JOURNALS LTD., LONDON, UK -----	11-13, 17

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 98/ 12236

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
Remark: Although claims 9 and 18, as far as in vivo methods are concerned, are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 18  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
Claim 18 relating to a method of modulating physiology or development of the cell using agonist and/or antagonist of C2, C2b, C18, C19 or C23 could only be searched partially, as the subject matter (agonist/antagonist) was not sufficiently disclosed.
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.